AD-A044 623 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/1

LARGE REGION LOCATION PROBLEMS.(U)

1977 D L LITWHILER

UNCLASSIFIED AFIT-CI-77-18 NL

1 OF 2

ADA044623

77-18

AD A044623

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

(1)

LARGE REGION LOCATION PROBLEMS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

DANIEL W LITWHILER, JR.

Norman, Oklahoma

1977

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT-CI-77-18
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER: 9 Doctoral Thesis
4. TITLE (and Subtitle): 6 Large Region Location Problems
5. TYPE OF REPORT & PERIOD COVERED: Dissertation
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 DANIEL L. LITWHILER, JR, CAPTAIN, USAF
8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT Student at University of Oklahoma, Norman, Oklahoma
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/CI, Wright-Patterson AFB OH 45433
12. REPORT DATE: 11 1977
13. NUMBER OF PAGES: 172 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 180p.
15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE AFR 190-17.

JERRAL F. GUESS, Captain, USAF
Director of Information, AFIT

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Attached

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

012200

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LARGE REGION LOCATION PROBLEMS

APPROVED BY

Adel A. Aly, Chairman

Marilyn J. Breen

Alan L. Dorris

Bobbie L. Foote

David C. Kay

DISSERTATION COMMITTEE

ABSTRACT

The study describes a number of algorithms for solving single facility deterministic location problems in which the planar assumption is not appropriate. Transformations on the non-Euclidean spherical space are combined with efficient solution techniques in E^n. Extensive use is made of projective, synthetic and analytic geometry.

Two algorithms are presented for solving single facility problems with the objective of minimizing the total sum of costs (minisum). Due to the non-convex nature of the problem, a local optimum is obtained. Computational experience in solving a number of test problems is reported. Theoretical results concerning narrowing of the search region are presented as well as a number of special properties of the problem. Application of the single facility results to the location-allocation class of multifacility minisum problem is investigated.

Algorithms for the single facility problem having the objective of minimizing the maximum distance (minimax) are also considered. It is seen that existing algorithms in E^2 can be used to solve the problem when all demand points are containable in a hemisphere. Application of single facility results to solution of a specific type of multifacility minimax problem is investigated.

ACKNOWLEDGEMENTS

Without the sponsorship of the United States Air Force this research effort would never have been possible. I would like to thank the USAF Academy and its Department of Mathematical Sciences for honoring me with the opportunity to return to school.

To Dr. Adel A. Aly, the chairman of my committee, I express my appreciation for his consistent support and expressions of confidence in my ability to perform up to his high standards.

To my excellent committee, including Drs. Marilyn J. Breen, Alan L. Dorris, and Bobbie L. Foote, I also express appreciation for their assistance and guidance throughout my doctoral program.

I especially wish to acknowledge committee member Dr. David C. Kay, whose expertise in geometry and efforts far beyond the "call of duty" were instrumental in developing the theoretical results concerning dominance in Chapter III and Appendix A.

Perhaps the greatest, but unseen, contributions were the sacrifices of my wife, Peggy, and children, Danny, Christopher, Kevin and Heather. For nearly three years they have managed to keep things moving smoothly (?) along while having only a part time husband and father. There must be special words to thank them, to describe what they have done, but they are all inherent in our family's living, together with love.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I.1 Background of the Research

Human concern about distance is an integral part of everyday life. Frequently decisions for or against an alternative are primarily based upon the constraining elements of "too near" or "too far." The theory of optimal facilities location provides tools which can assist one in selecting sites for servicing facilities that will directly affect the distance traveled.

Cooper (1963) credits Cavalieri with first considering, in 1647, the problem of finding a point at which the sum of those distances from three given points is a minimum. This is referred to variously as the Steiner or Fermat problem in the literature. Sylvester (1857) evidently first proposed the problem of finding the points in the plane such that the maximum distance to a finite set of points in the plane is minimized. Although location theory has developed far beyond these fundamental mathematical problems, new observations and results are still being reported concerning them (Sokolowsky 1976).

A new era in location theory began with development of an iterative process to solve a generalization of the Steiner-Fermat problem. The technique, developed independently by Cooper (1963) and Kuhn and Kuenne (1962), was first proposed by Weiszfeld (1936). In the past twenty years literally hundreds of papers have appeared on the subject in many different fields. Regarding the general literature of location theory as it exists today, ReVelle et al. (1970) found it convenient to consider two main structural categories:

1. Location on a plane, characterized by an infinite solution space and a distance measurement according to any particular metric.

2. Location on a network, characterized by a solution space consisting of points on a network and the distance or time measurement d_{ij} is the length/time of the shortest path from node i to node j.

This morphology is generally accepted as representative of location systems theory.

Early maps of the world (circa eleventh century) depicted the earth as a flat plane. The discovery that the earth was essentially spherical created problems for cartographers that are still being investigated today. It is well known (Hilbert 1952) that there exists no isometric (length-preserving) transformation from a sphere to the plane. This property, or lack thereof, eliminates the possibility of a transformation so that Euclidean geometry can be used, in the large, to measure spherical distances.

Due to the relative intractability of working in non-Euclidean space, however, much effort has been expended in the past in attempting to develop transformations from non-Euclidean to Euclidean space which preserve desired properties. A review of such attempts is provided by Angel and Hyman (1972). Also, in another vein, research is ongoing in attempts to create optimal transformations from the sphere to the plane which minimize the error of a given parameter (area, distance, etc.) for a specific region. Recent work has been done by Milnor (1969) and Gilbert (1974).

In spite of the foregoing efforts there remain problems in location theory for which the Euclidean assumption is totally inappropriate. Unfortunately, though, the majority of theoretical developments in several related fields (location theory, theoretical geography, and regional science) require a geometrical framework based upon the assumption of a Euclidean plane. Considering the nature of the earth, it is surprising that so little consideration has been given to non-Euclidean bases, specifically the great circle metric. It is such a situation to which this research is addressed.

A simple illustration of the magnitude of error which can occur is provided in Table 1.1 for a three point Steiner problem on the sphere (see Data Set D9 of Appendix B). Using both a Euclidean assumption and a spherical assumption, the problem was solved by an existing algorithm for the plane and methods developed in this thesis for the sphere. The coordinates of the points are in degrees latitude and longitude. For the planar problem, the spherical coordinates were used as Cartesian coordinates.

	Location (Lat., Long.)	Objective Function Value	
		Great Circle Metric	Euclidean Norm
Optimum on Sphere	(80.6,75)	2.08187	2.7728
Optimum on Plane	(61.75,75)	2.15123	2.70985

Table 1.1. Sphere vs. plane, a Steiner problem.

When using the planar assumption there is an 18.85^{o} error in location of the optimum facility, or approximately 1300 miles on the earth's surface. Also, there is over a 30% error in the optimum objective function value.

Due to the increasingly global nature of corporations and international defense structures, it would seem useful to consider large region location problems. As Warntz (1966) declares, perhaps the time of "community earth" is not too far removed into the future. Due to the infeasibility of developing a transformation to handle large region problems on a sphere via Euclidean geometry, there is a need to develop solution techniques for such problems.

It is recognized that the earth is a spheroid rather than a sphere. Computation for the spheroid is tedious at best, but the error due to a spherical assumption is not too significant. Use of spheroidal equations would only be warranted where extreme accuracy is sought, and when all other aspects of the problem are handled with utmost care. If the earth were represented by a spheroid with an equatorial diameter of 25 feet, the polar diameter would be approximately 24 feet 11 inches (Deetz and Adams 1945). Maling (1973) cites a study in which the average distance error when using a spherical assumption is less than 1.0% between 200 points in the United States.

1.2 Location Problems Considered

This study deals with optimizing the point location of a single servicing facility in a three dimensional, constrained, continuous solution space for a finite system of known fixed demand points located on a sphere. The demand points are connected with the servicing facility by communications/transportation links. The demand points cover a large region, i.e., a region in which the planar assumption introduces considerable error. A single equality constraint forces the solution to be two dimensional in the three dimensional space by restricting solutions to be on the surface of a sphere. That is, the communications/transportation links between the servicing facility and the demand points are via great circle arcs on the surface of the sphere which contains the demand points. Costs due to transportation are a linear function of distance. Except for specific examples, a unit sphere shall be assumed. This is done without loss of generality since geodesic distance, s, is directly related to the radius, r, via $s = r\theta$.

Two different objectives will be considered:

i. Minisum. Minimize the total sum of costs. Having M existing facilities located at known distinct points $P_1, P_2, \ldots, P_M$; a new facility is to be located at point X. Costs of a transportation nature are incurred that are directly proportional to the distance metric between the new facility and existing demand point.

Definition 1.1 Given any two points $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ in Euclidean E^n space, and $p > 1$, the ℓ_p metric between X and Y is:

$$|X-Y|_{\ell_p} = [\sum_{i=1}^{n} |x_i - y_i|^p]^{1/p}$$

(Note that $p = 1$ and $p = 2$ represent the rectilinear and Euclidean norms, respectively.)

If W_i is the product of cost per unit distance and number of trips per time period between X and P_i, the total cost per period is given by:

$$f(X) = \sum_{i=1}^{M} W_i |X-P_i|_{\ell_p}$$

The single facility location problem using the minisum objective is to determine the location of the servicing facility, say X*, that minimizes f(X), the period's total transportation cost.

For the large region location problem one can define $\rho(X,P_i)$ as the shortest great circle distance between demand point P_i and servicing facility X. This measurement of distance is shown to be a metric by Blumenthal (1961). The problem then becomes:

$$\underset{X}{\text{Minimize}} \quad \sum_{i=1}^{M} W_i \rho(P_i,X)$$

$$\text{Subject to } |X|_{\ell_2} = 1, \quad X \varepsilon E^3$$

where $P_i \varepsilon E^3$ are points on the unit sphere for $i = 1,\ldots,M$.

ii. Minimax. Minimize the maximum distance. That is:

$$\underset{X,Z}{\text{Minimize}} \ Z$$

$$\text{Subject to } |X-P_i|_{\ell_p} \leq Z \quad i = 1,\ldots,M$$

where Z is (geometrically) the radius of a sphere centered at X, and P_i and X are as above.

Suppose a known finite number of points on the surface of a sphere is given and it is desired to locate a single point on the sphere's surface so that the maximum great circle distance between this point and the given point is minimized. The problem is then formulated as:

$$\begin{aligned} &\underset{X}{\text{Minimize}} \quad Z \\ &\text{Subject to } \rho(X,P_i) \le Z \\ &\qquad\qquad |X|_{\ell_2} = 1 \end{aligned}$$

where X and P_i are as above.

I.3 Application of Research

This research has application in any large region, long range single facility minimax or minisum location problem and, as will be pointed out in Chapters III and IV, may be useful in development of techniques to handle the counterpart multifacility problem. Such topics as detection station placement, naval deployment, location of international headquarters or distribution/marketing centers, and location of long range weapons systems fall within the purview of this research. It would seem especially pertinent in the area of long range communications. The importance in radio engineering stems from the fact that radio transmissions follow a great circle track.

I.4 Scope and Limitations

The research is limited to consideration of the single facility location problem using a great circle metric. The servicing facility is restricted to the spherical surface upon which a finite number of "weighted" demand points are located. As cited previously, except for the restriction on location of the servicing facility, the problem is unconstrained. Two criteria, minimization of total costs and minimization of maximum costs, are investigated.

Extensive use of synthetic, projective and analytic geometry is made throughout the research in development of solution procedures and establishing the theoretical results.

There is a definite gap in the theory of location with regard to the great circle metric. This research is motivated by recognition that there are situations in which it is necessary to use special techniques, notably when the regions considered are larger than a hemisphere. The research investigates the properties of such problems and develops approaches for handling them. It is hoped this work will ultimately stimulate interest in the development of efficient approaches for the multifacility location problem on the sphere.

I.5 Order of Presentation

Due to the magnitude of work that has been done on the generalized Weber problem and other location models, Chapter II contains a survey of only the literature directly related to the research effort. Extensive bibliographies on the general literature are referenced. Chapter III presents the research findings concerning the minisum single facility problem using a great circle metric. Two heuristic algorithms are developed for solution of the problem. In Chapter IV the related minimax problem is examined. It is shown that existing algorithms for the problem in E^2 and E^3 can be adapted to solve the large region problem in certain cases. Chapter V presents computational experience with the algorithms developed in Chapter III. A number of example problems are solved and results discussed. The research is summarized and recommendations for future research are made in Chapter VI.

CHAPTER II

STATE OF THE ART

This chapter presents a review of previous research pertinent only to location problems in large regions. As mentioned in Chapter I, two specific types of objective functions are of interest:

1. *Minimization of total costs* (Minisum), and
2. Minimization of maximum distance (Minimax).

Concerning the general literature on location, one may refer to extensive bibliographies (Francis and Goldstein 1974; Lea 1973), and a textbook (Francis and White 1974).

Only recently have researchers shown any interest concerning location problems of the minisum variety for regions so large that a Euclidean (planar) assumption is not appropriate. Wendell (1971) provided a brief discussion of the minisum single facility location problem on the earth's surface. An approximation technique was given which used Schwartz's inequality to find an explicit approximate solution (see Section III.8). Methods for obtaining an exact solution were not considered.

Lea (1973) states that some work has been done on the problem by Krolikowski. The Newton-Raphson iteration method was evidently used to solve the problem and contouring was proposed both to supplement problem solution and to overcome geographic infeasibilities.

Katz and Cooper (1975, 1976) discussed the theoretical and computational aspects of the problem, with some attention paid to other metrics on the sphere. An interactive method employing a normalized gradient and an acceleration scheme due to Steffensen (Henrici 1964) is used for finding a local optimum. Steffensen's technique is a standard method for accelerating convergence in an iterative algorithm.

Nothing in the literature indicates a complete or explicit work related to the subsequent research effort. Nearly all the published work that refers to location in a large region or on a spherical surface has been concerned with the minimax criterion. However, most of the effort in the past has been directed toward solving the problem for the continuous case, i.e., locating n facilities on the sphere so that the maximum distance to <u>any</u> point on the sphere is minimized.

The attempts at solving this problem, and a concise statement of it, are given by Tóth (1973). An excellent review of solutions for the cases n = 5 and n = 7 is found in Meschkowski (1966). There are only a few values of n for which the solution is known and for arbitrary values of n there is not even a reasonable conjecture concerning the solution.

Concerning the minimax criterion when only a finite number of demand points exist on the sphere (the problem of interest), evidently nothing has been done in over a century. The problem in E^2 was first proposed by Sylvester (1857). As cited by Sylvester (1860), the problem was solved for E^2 by Peirce. The technique was rediscovered by Chrystal (1885). A more recent presentation of the approach is given by Rademacher and Toeplitz (1957) in a book of mathematical diversions. In Sylvester's paper the problem was erroneously claimed to be completely analogous to the one on the sphere. This claim is discussed further in Section IV.4.

Peirce's rudimentary technique for E^2 is suitable for solving the problem by hand. Recently Elzinga and Hearn (1972a) presented a similar algorithm which was conducive to programming on a computer and hence is more efficient. No such efficient technique exists for the counterpart problem on the sphere.

Since little has been done concerning the large region location problems, a Euclidean assumption was made, using latitude and longitude as Cartesian coordinates. Any large region "real world" example problem in the literature is invariably concerned with not-so-large a region situated in the middle latitudes, thus keeping the error due to a Euclidean assumption within bounds of acceptability. Notable examples of such problems are presented by Kuhn and Kuenne (1962), Smallwood (1965), and Chapelle (1969). Two of these problems will be discussed in more detail in Chapter V.

Kuenne and Soland (1971, 1972), in a study of the multifacility generalized Weber problem, convert latitude and longitude to coordinates on a Mercator Projection (Warntz and Wolff 1971), compute all distances as rhumb-line map distance, and convert to approximate great circle distances. Although it worked well for their purposes, the error in this method increases dramatically as either polar region is approached or regions larger than a hemisphere are considered. Their approach is discussed further in Section III.14.2.

Considering the foregoing, it is evident that researchers were aware of the problems encountered in using a Euclidean assumption when demand points are scattered over a large region, but to date little has been done to resolve them.

CHAPTER III

MINISUM SINGLE FACILITY LOCATION PROBLEMS

III.1 Introduction

In this chapter minimization of the total costs for a single facility location problem on a sphere using the great circle metric is addressed. Two new algorithms for solution of the "minisum" problem are developed, and a number of properties of the problem are presented.

As stated in Chapter I, a general form of the problem for ℓ_p metrics is given by:

$$\min_X f(X) = \sum_{i=1}^{M} W_i |X-P_i|_{\ell_p}$$

where: P_i for i=1,...,M are fixed demand points

W_i for i=1,...,M are non-negative weights, and

X is the unknown location of the servicing facility, $X \varepsilon E^2$

The objective, then, is to determine the location of the new facility, say X^*, that minimizes f(X), the total "transportation" cost. Cost is considered to be strictly a function of great circle distance and demand point weights.

III.2 Problem Formulation

On the sphere, formulation of the problem requires use of inverse trigonometric functions. There are a number of formulations, each having its own advantages and disadvantages. In general, the problem can be formulated as

$$\min_{X} \sum_{i=1}^{M} W_i \rho(P_i, X) \qquad (3.2.1)$$

$$\text{Subject to } |X|_{\ell_2} = 1, \ X \varepsilon E^3$$

where: $\rho(P_i, X)$ is the great circle distance between demand point P_i and servicing facility X, and $|\cdot|_{\ell_2}$ is the Euclidean norm. Without loss of generality, the problem is formulated for the unit sphere since arc length on a spherical surface is directly proportional to the radius.

Recognizing that the shortest distance between two points on a sphere is via the shorter great circle arc connecting them (Lyusternik 1964) it can be shown that

$$\rho(P_i, X) = 2 \text{ Arcsin } \frac{|P_i - X|_{\ell_2}}{2} \qquad (3.2.2)$$

Thus, one has:

$$\min_{X} \sum_{i=1}^{M} 2W_i \text{Arcsin } \frac{|P_i - X|_{\ell_2}}{2} \qquad (3.2.3)$$

$$\text{Subject to } |X|_{\ell_2} = 1, \ X \varepsilon E^3$$

where: W_i, P_i, and X are defined as in Section III.1.

It is easily seen that the argument for Arcsin is restricted between 0 and 1, and that the function is convex in this region. It is well-known that a convex function operating on a convex function is likewise convex. Recognizing that the argument is convex and that the objective function is a non-negative sum of convex functions it follows that it, too, is convex. Unfortunately, though, the solution space is not convex, resulting in a non-convex programming problem. This is demonstrated graphically in Figure 3.1. Although this form of the problem is revealing with regard to properties of the unconstrained problem, it was found to be not as efficient computationally in repeated calculations of objective function values as the next formulation (about 20% slower on an IBM 370/158J).

A basic result of elementary calculus is that if two lines L_1 and L_2 have direction cosines (λ_1,μ_1,ν_1) and (λ_2,μ_2,ν_2), and if θ is the angle between L_1 and L_2 then $\cos\theta = \lambda_1\lambda_2 + \mu_1\mu_2 + \nu_1\nu_2$. Defining ψ_i as the angle between the rays in E^3 determined by $\overrightarrow{OP_i}$ and $\overrightarrow{OX}$, where O is the center of the sphere, and observing that minimizing the weighted sum of the arcs (great circle distances) is equivalent to minimizing the weighted sum of the angles which subtend the arcs, the objective is

$$\underset{X}{\text{Min}} \sum_{i=1}^{M} W_i\psi_i \tag{3.2.4}$$

$$\text{Subject to } |X|_{\ell_2} = 1$$

On a sphere of radius 1 with center at the origin in E^3, demand points $P_i = (a_i,b_i,c_i)$, i=1,...,M and

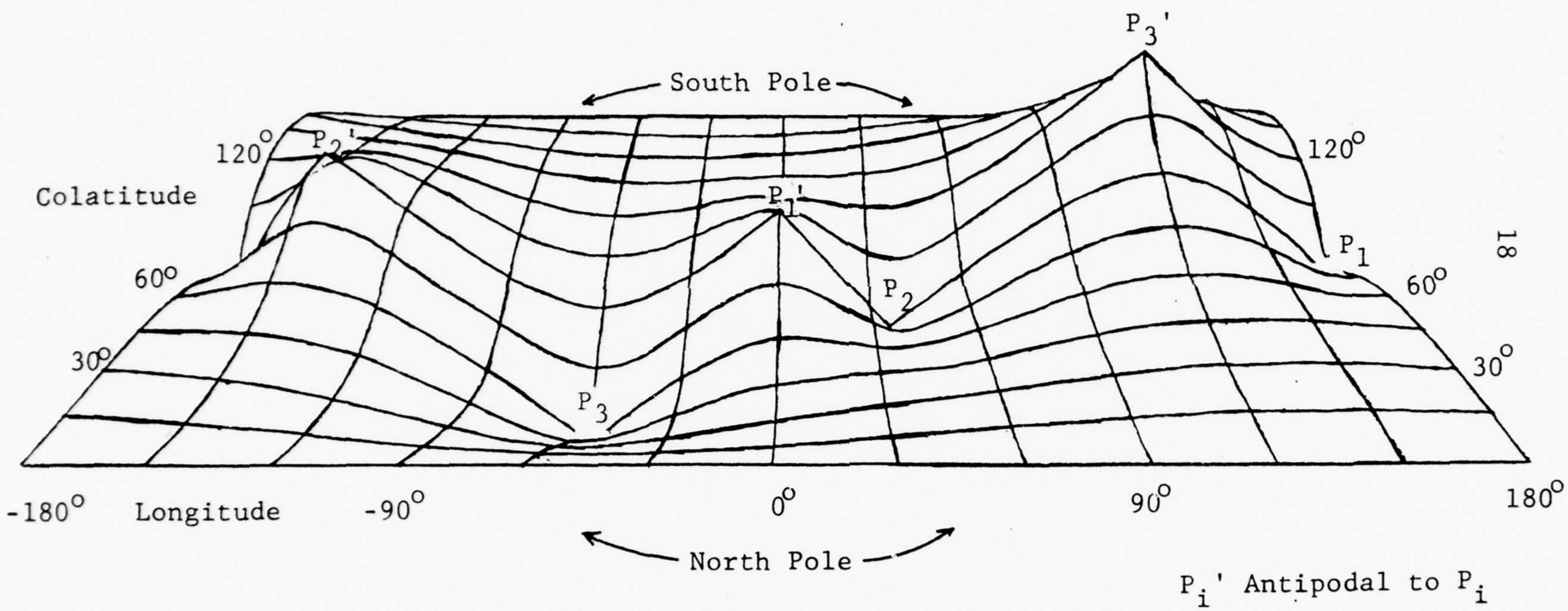


Figure 3.1. 3-D Minisum function plot; Data Set D1, Appendix B.

$X = (x_1, x_2, x_3)$ it follows that:

$$\psi_i = \text{Arccos}\ (a_i x_1 + b_i x_2 + c_i x_3) \qquad (3.2.5)$$

where: $\|X\|_{\ell_2} = 1$. Converting to spherical coordinates one gets the following formulation:

$$\text{Min} \sum_{i=1}^{M} W_i \ \text{Arccos}\ (a_i \sin\phi\cos\lambda + b_i \sin\phi\sin\lambda + c_i \cos\phi) \qquad (3.2.6)$$

$$\text{Subject to } -\pi < \lambda \leq \pi$$

$$0 \leq \phi \leq \pi$$

where: $x_1 = \sin\phi\cos\lambda \qquad W_i \geq 0 \quad i=1,\ldots,M$

$x_2 = \sin\phi\sin\lambda$

$x_3 = \cos\phi$

This formulation, equivalent to (3.2.3), was found to be efficient for the two algorithms developed in this chapter. Using formulation (3.2.6), though, an essentially unconstrained problem with two bounded variables is solved. The bounds are necessary only because of periodicity of the objective function. A conspicuous disadvantage, which could affect its use in some algorithms, is that one knows little about the properties of the objective function. The Arccos function is neither convex nor concave over the domain.

Any point on the sphere can be identified by a two-tuple (ϕ, λ) where ϕ is the colatitude $(0 \leq \lambda \leq \pi)$ and λ is the longitude or meridian $(-\pi < \lambda \leq \pi)$. Letting $P_i = (\phi_i, \lambda_i)$ and $X = (\phi, \lambda)$, via the Law of Cosines for Elliptic Geometry

(Kay 1969) and properties of the cosine function it follows that:

$$\cos[\rho(P_i,X)] = \cos\phi_i\cos\phi + \sin\phi_i\sin\phi\cos(\lambda-\lambda_i) = A_i$$

and thus:

$$\rho(P_i,X) = \text{Arccos}\ (A_i)$$

The goal, then, is to:

$$\text{minimize} \sum_{i=1}^{M} W_i\ \text{Arccos}\ (A_i) \qquad (3.2.7)$$

$$\text{Subject to} \quad 0 \leq \phi \leq \pi$$

$$-\pi < \lambda \leq \pi$$

This formulation, although similar to the previous one, was found to be computationally inefficient. This is primarily due to time required for Taylor Series approximations of trigonometric functions on a digital computer. Its principal advantages are that one works directly with spherical coordinates, along with the property that the formulation is essentially an unconstrained problem.

III.3 Fundamental Properties

In this section the non-convexity of the problem is demonstrated, along with non-differentiability of the objective function and the fact that the domain of the objective function is restricted.

III.3.1 Non-convexity

The unfortunate characteristic of non-convexity, which is not a factor in the planar single facility location problem, is best seen through an example. Consider the three equally weighted demand points in Data Set D1 of Appendix B and suppose the objective is to minimize the total sum of great circle distances. A three dimensional graph of the objective function values as a function of ϕ and λ is seen in Figure 3.1. It is depicted as a flattened sphere with infinite distortion along the $\phi\lambda$ plane at the North and South Poles. It is clear from the graph that local minima exist at P_1, P_2, and P_3, with possible doubts as to what is happening behind the peak at P_3'. It turns out that all the local minima are visible, and it will be seen in Chapter V that the global minimum occurs at P_3.

Figure 3.2, a contour plot of the minisum objective function using Data Set D2 of Appendix B, exhibits the possible existence of alternate optima on the sphere, namely the entire arc P_2P_3. The graph of the objective function over the great circle arc as a function of longitude is piecewise linear.

III.3.2 Non-differentiability

Besides the non-convexity of the problem, difficulties arise as to differentiability of the objective function, necessitating consideration of techniques to circumvent these difficulties during any search procedure. Although non-differentiability occurs only at the demand points for the counterpart planar Euclidean norm problem, on the sphere the objective function is non-differentiable at both

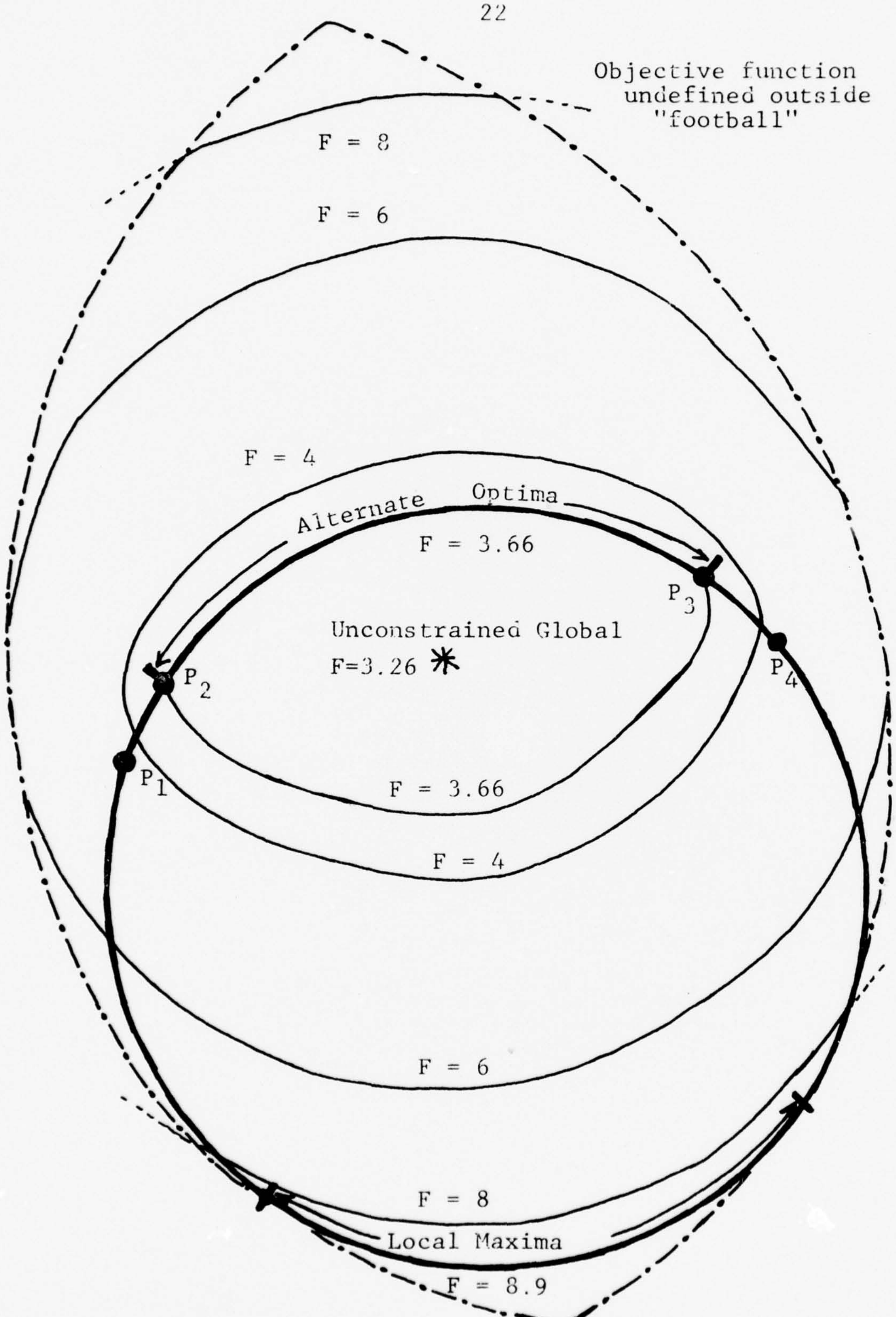


Figure 3.2. Contour plot; Data Set D2 of Appendix B.

the demand points and at the unique points antipodal ($\rho(x,y) = \pi$) to each of the demand points. Referring to Figure 3.1 again, the "knife-edged surface" described by Vergin and Rogers (1967) appears at these points. The reason for this occurrence can be seen by examining the derivative of Arccos X:

$$\frac{\partial \text{Arccos } X}{\partial X} = -1/\sqrt{1-X^2}$$

Note that the first derivative is undefined when $X = \pm 1$, which occurs for arcs of length 0 and π, which in turn correspond to the situations when one is at a demand point or its antipodal point.

Concerning the search techniques which will be developed, one of them requires consideration of this property since it employs a gradient search. The other algorithm is derivative-free.

IV.3.3 Restricted Domain of Objective Function

Another property of the single facility minisum sphere problem not encountered in the planar case is that the domain for the objective function is restricted. This is because the real argument Z of the Arccos function may not exceed the bound $|Z| \leq 1$. Given that X is a possible servicing facility location, certainly the objective function is defined for all possible X contained within or on the unit sphere, but not defined for all X outside the sphere. In Figure 3.2 this is demonstrated for a four point problem in the degenerate case of location on a great circle arc. The "outer" area is not within the domain of definition. As the number of demand points increase around the sphere, note that the domain of definition will shrink to approximate the sphere.

This restricted domain of definition must be considered when using any search technique calling for projections. Care must be taken to insure that the search does not leave the domain of definition, thus causing premature termination.

III.4 Dominance Properties on the Sphere

III.4.1 Introduction

Although the general minisum problem on the sphere has many undesirable properties, there are situations in which one can reduce the search region for a global optimum. For example, it seems plausible and intuitive that if all the demand points are located on a hemisphere, then any search for an optimal solution can be restricted to this region. Using specific results of convexity theory for spherical geometry, this intuitive concept is now generalized to demonstrate that any search for an optimal solution to (3.2.1) can be restricted to the spherically convex hull of the demand points.

The major results of this section are based upon a generalization of Kuhn's (1967) concept of dominance due to Wendell and Hurter (1973). It is described by:

Definition 3.4.1 A point x' dominates a point x with respect to $P_1, \ldots, P_i, \ldots, P_M$, demand points, if and only if $\rho(x', P_i) \le \rho(x, P_i)$ for all i.

As an immediate result, if x' dominates x with respect to P_i for all i, and W_i is a non-negative constant for all i, x' has the property that:

$$\sum_{i=1}^{M} W_i \rho(x', P_i) \le \sum_{i=1}^{M} W_i \rho(x, P_i)$$

So, by showing for any $x \notin V$, where V is a spherically convex hull, that there exists an $x' \varepsilon V$ such that x' dominates x one is assured of the existence of an optimal solution x* such that x* is in V.

Prior to establishing the main results it is necessary to introduce some concepts and lemmas. Terminology and proofs of supportive lemmas are in Appendix A.

III.4.2 Dominance within Spherically Convex Hull

In this section it is established that any search for an optimal solution to (3.2.1) can be restricted to the spherically convex hull containing the demand points.

Theorem 3.4.1 Given a set of demand points $\{P_i | i=1, \ldots, M\}$ (not consisting of two antipodal points) whose convex hull is $V = \text{conv}\{P_i | i=1, \ldots, M\}$. For any $x \varepsilon S^2$ such that $x \notin V$, there exists $x^* \varepsilon V \subset S^2$ such that x* dominates x in the great circle metric with respect to P_i, $i=1, \ldots, M$.

<u>Proof</u>: By Lemma A.8, V is closed. So applying Lemma A.4, consider the support line L orthogonal to line $\overleftrightarrow{xp}$ at p. The proof is in two parts.

<u>Part A</u>. Suppose that V lies entirely on a line through x, and further that the measure of V = $\overline{pq}$ is $\leq \pi$ (note that if the measure is $> \pi$, then it is equal to 2π). For $x \notin V$, designate p as the closest point of V to x. If $\rho(x,p) \geq \pi/2$, choose x^* as the midpoint of $\overline{pq}$. Since $\rho(p,q) \leq \pi$, then $\rho(x^*,P_i) \leq \pi/2$ for all $P_i \subset V$, i=1,...,M; and x^* dominates x.

If $\rho(x,p) < \pi/2$, consider the following. If $\rho(p,q) \leq \rho(x,p)$ then the choice $x^* = p$ obviously dominates x since for any demand point P_i, (xpP_iq) holds and $\rho(x^*,P_i) = \rho(p,p_i) \leq \rho(p,q) \leq \rho(x,p) \leq \rho(x,P_i)$. Otherwise, take $x^* \varepsilon \overline{pq}$ such that (xpx^*) and $\rho(x,p) = \rho(p,x^*)$. Also, define x' and p' as the antipodal opposite points of x and p, respectively. Let P_i be a demand point.

Case I. $P_i \varepsilon \overline{px^*}$ (Figure 3.3a)

Then (xpP_ix^*) holds and $\rho(x^*,P_i) \leq \rho(x^*,p) = \rho(x,p) \leq \rho(x,P_i)$ by properties of betweenness; hence x^* dominates x.

Case II. $P_i \varepsilon \overline{x^*x'}$ (Figure 3.3b)

Then (xpx^*P_i) holds and $\rho(x,P_i) = \rho(x,x^*) + \rho(x^*,P_i) \geq \rho(x^*,P_i)$ by properties of betweenness, and dominance follows.

Case III. $P_i \varepsilon \overline{x'p'}$ (Figure 3.3c)

Then $(xp'P_i)$ holds, as does $(x^*x'P_ip')$. Now since,

$$\rho(p,p') = \pi = \rho(p,x) + \rho(x,p')$$

$$= \rho(p,x^*) + \rho(x^*,p')$$

$$= \rho(p,x) + \rho(x^*,p'),$$

it follows that $\rho(x,p') = \rho(p',x^*)$.

Adding $\rho(p',P_i)$ to both sides, and utilizing the fact that $(xp'P_i)$, one has

$$\rho(x,P_i) = \rho(p',x^*) + \rho(p',P_i) \geq \rho(p',x^*)$$

Since (x^*P_ip'), it follows from betweenness properties that

$$\rho(p',x^*) = \rho(x^*,P_i) + \rho(P_i,p') \geq \rho(x^*,P_i)$$

Thus $\rho(x,P_i) \geq \rho(x^*,P_i)$ and x^* dominates x.

Since all possible cases have been considered, the proof of Part A is complete.

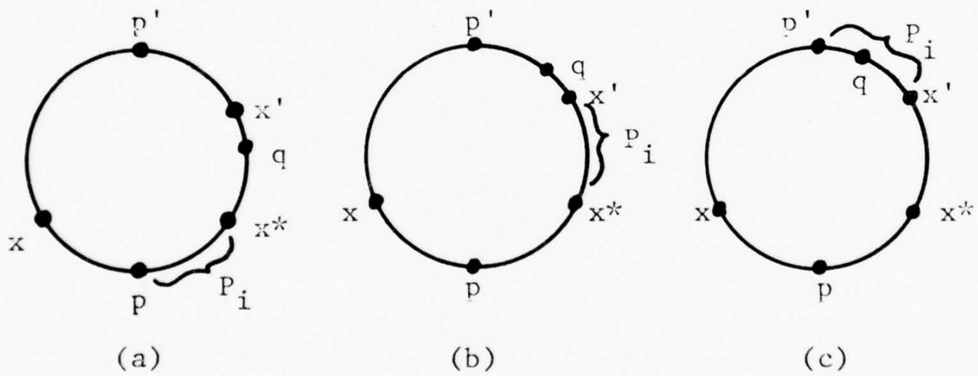


Figure 3.3. Theorem 3.4.1, Part A.

Part B. Next, assume V does not lie on a line through x. By Lemma A.5, there exist exactly two distinct rays $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$ with b_1, b_2 ϵ bd V such that the lines $\overleftrightarrow{xb_1}$ and $\overleftrightarrow{xb_2}$ are support lines of V and each ray $\overrightarrow{xv}$ for $v \epsilon V$ either coincides with $\overrightarrow{xb_1}$, or $\overrightarrow{xb_2}$, or lies between $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$. In particular, $\overrightarrow{xp}$ is between $\overrightarrow{xb_1}$, $\overrightarrow{xb_2}$ or coincides with them, so the sum of the angles $\measuredangle pxb_1$ and $\measuredangle pxb_2$ is no greater than π, and at least one of them, consequently, can be no greater than $\pi/2$. We may assume without loss of generality $\measuredangle pxb_1 \leq \pi/2$. Now either $\measuredangle pxb_2 \leq \pi/2$ or $\measuredangle pxb_2 > \pi/2$. The cases are considered separately.

Case I. Suppose $\measuredangle pxb_2 \leq \pi/2$ (see Figure 3.4a). Let $x^* = p$, and suppose v is any point of V not on line $\overleftrightarrow{xp}$ (there must exist at least one such point under the assumption). If v is on the b_k-side of $\overleftrightarrow{xp}$ $(k = 1,2)$ then since $\overrightarrow{xv}$ coincides with $\overrightarrow{xb_k}$ or lies between $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$, $\measuredangle pxv \leq \measuredangle b_k xp \leq \pi/2$. On the other hand, with v on L or on the opposite side of L as x, one of the opposing rays from p on L either coincides with $\overrightarrow{pv}$ or lies between $\overrightarrow{px}$ and $\overrightarrow{pv}$ so that $\measuredangle xpv \geq \pi/2$. Hence in all cases $\measuredangle pxv \leq \measuredangle xpv$. If equality holds, $\rho(p,v) = \rho(x,v)$. If inequality holds, using the properties that the largest side of a spherical triangle is opposite the largest angle (Kay 1969, Theorem 31.5), it follows that $\rho(p,v) < \rho(x,v)$. That is, $\rho(x^*,v) \leq \rho(x,v)$ for all v in V not on line $\overleftrightarrow{xp}$. However, by connectedness of V this holds for all points of V. So, for all demand points P_i it follows that $\rho(x^*,P_i) \leq \rho(x,P_i)$. Dominance of $x^* \epsilon V$ has thereby been proved for this case.

Case II. Suppose $\measuredangle pxb_2 > \pi/2$. According to Lemma A.6 one may choose $u \epsilon \overline{xb_2}$ and $x' \epsilon \mathrm{bd}\, V$ such that $\rho(x,u) = \rho(u,x')$ and $\overleftrightarrow{ux'}$ is a line of support of V. (Figure 3.4b). In this case put $x^* = x'$. Let u' be the antipodal point of u.

Then both triangles xux' and xu'x' are isosceles. Once again $\overrightarrow{xx'}$ either coincides with $\overrightarrow{xb_k}$ or falls between $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$. Let v be any point of V not on $\overleftrightarrow{xx'}$; then $\overrightarrow{xv}$ coincides with $\overrightarrow{xb_k}$, k = 1,2, or lies between them. In the case when $\overrightarrow{xv}$ lies on $\overrightarrow{xb_2}$ or between $\overrightarrow{xx'}$ and $\overrightarrow{xb_2}$, ray $\overrightarrow{x'v}$ lies between $\overrightarrow{x'u}$ and $\overrightarrow{x'y}$ (y is any point on $\overleftrightarrow{xx'}$ such that (xx'y)) or coincides with $\overrightarrow{x'u}$. Hence $\measuredangle xx'v \geq \measuredangle xx'u = \measuredangle x'xu \geq \measuredangle vxx'$ and by properties of triangles having unequal angles, $\rho(x',v) \leq \rho(x,v)$. The same result holds when $\overrightarrow{xv}$ lies on the other side of line $\overleftrightarrow{xx'}$, appealing to equal angles $\measuredangle u'xx'$ and $\measuredangle u'x'x$. Again, as in Case I, $\rho(x^*,v) \leq \rho(x,v)$ holds for all $v \in V$ and specifically for all demand points $P_i = v \in V$, proving dominance of x^* in this last case.

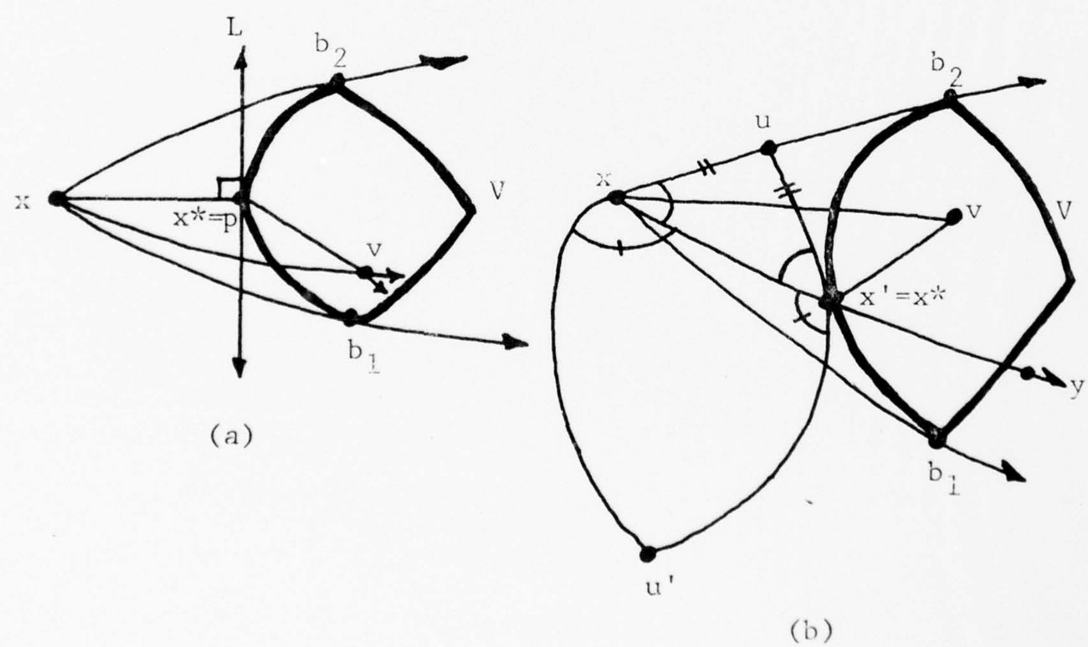


Figure 3.4. Theorem 3.4.1, Part B.

Corollary 3.4.1 For the great circle metric, there exists an optimal solution x^* to the problem (3.2.1) such that $x^* \varepsilon \text{ conv } \{P_i | i=1,\ldots,M\} = V$.

Proof: By using a spherical coordinate system the constraint is satisfied implicitly. The result follows from the observation that if x_0 dominates x, then

$$\sum_{i=1}^{M} W_i \rho(x_0, P_i) \leq \sum_{i=1}^{M} W_i \rho(x, P_i).$$

One need only consider elements within the region since if there exists x_0 outside of V which is optimal, Theorem 3.4.1 guarantees the existence of an x^* within V which dominates x_0.

III.4.3 Determining Demand Points' Hull Characteristics

In order to apply the foregoing results, it is of course necessary to insure that the set of demand points can be contained in a hemisphere. For if this is not possible, the spherically convex hull is the entire sphere, and the search region is not reduced at all. The logical procedure is to coordinatize, plot the points on a sphere and visually verify whether the set is containable in a hemisphere.

Blumenthal (1956) defined a global subset as a subset G of S^2 which is not contained in any hemisphere of S^2. For the case of m=4 points, he established necessary and sufficient conditions for the points to form a global subset. Clearly for $m \leq 3$, the points can be contained in a hemisphere. Nothing is known about the general case of $m > 4$.

What follows is a rudimentary heuristic procedure for determining whether the set $\{P_i | i=1,...,m\}$ forms a global subset. First, rank all inter-demand point distances from the smallest to the largest and select the most distant pair. Using either of the points, say P_k, transfer the pole to that point. Rank the longitudes of all the remaining demand points in ascending order, and see if any absolute difference between the longitudes of "adjacent" demand points (with reference to rank order of longitudes) is greater than or equal to π. If so, one of the two hemispheres determined by the line through P_k and the point having minimum of the two longitudes contains the set of demand points. It is the hemisphere which contains the point with the maximum of the two longitudes. The absolute difference between the minimum and maximum longitude of all demand points (other than P_k) is given by 2π - (max longitude - min longitude).

If no absolute difference is greater than π, choose the other point of the original pair, unless it is an antipodal point, and repeat the procedure. This is continued through the list of inter-demand point distances until either an enclosing hemisphere is found or all points have been considered.

III.4.4 Summary

In this section it was shown that the search for an optimal solution can be restricted to the convex hull of the demand points. Note that if the set of demand points is not containable in a hemisphere, the convex hull is the entire sphere, and no reduction in the search area is possible. The only sets of demand points of interest, then, are those containable in a hemisphere.

A rudimentary procedure is provided to determine whether the spherically convex hull is containable in a hemisphere for a particular set of demand points. Certainly, based upon these results, one would not pick starting solutions outside the convex hull when applying an iterative technique.

IV.5 Analogue Models

Mechanical and electrical analogue models for the Euclidean norm single facility location problem have been devised and successfully employed. The purpose of this section is to extend these models to handle the related problem on S^2.

III.5.1 Mechanical Analogue

A description of the basic analogue model is given by Eilon et al. (1971), Francis and White (1974), Lyusternik (1964), and Haley (1962), among others. Utilizing strings and weights, it was introduced by Georg Pick in the early

1900's in Alfred Weber's Über den Standort der Industrien (1909). Description of a case study using this model is given by Burstall et al. (1962).

The model which follows will solve the single facility great circle metric problem (3.2.1). It can be used for regions as large as a hemisphere without difficulty.

First, a highly polished sphere (to minimize friction effect) is coordinatized and demand points are plotted. Holes are then drilled at the demand point locations and strings are passed through the holes with the ends tied to a small ring resting on the exterior surface of the sphere. The other ends are passed through a stationary ring at the sphere's center. Weights proportional to the respective demands at each point are tied to the appropriate strings (see Figure 3.5).

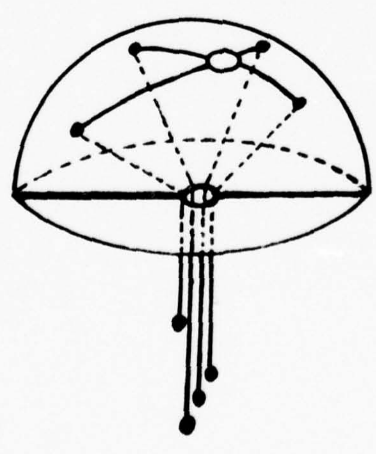


Figure 3.5. A mechanical model.

The outer ring (on frictionless bearings) is pulled to one side and then released. The weights will pull the ring to a point of minimum potential energy at which there is a local minimum objective function value. Due to the demonstrated non-convexity of the problem, it is possible, although unlikely, that the ring could start at a local maximum and not move. For this reason, it should be moved slightly to one side to see if it returns to the same point. Local maxima and minima demonstrate points of unstable and stable equilibria, respectively. The unstable equilibria occur due to the duplicity of paths between any two points; that is, the long and short great circle arcs. This property is treated in detail by Lyusternik (1964).

In regions larger than a hemisphere physical difficulties can arise in the model as the ring approaches a point y antipodal to a demand point x ($\rho(x,y) = \pi$). At such a point there are an infinite number of paths of equal length to the antipodal demand point. Yet, by moving an arbitrarily small distance in any direction, the shortest path becomes unique. The difficulty arises in the physical slipping of the string to a position 180° opposite to the existing position just prior to the ring's passing through the antipodal point. Friction and interference from other strings would prevent free movement of the ring in such a case.

The advantage of this model is that rough estimates of location for local optima can be found visually rather quickly by starting the ring at various places around the sphere, thus providing good starting solutions for any iterative technique.

Disadvantages include the friction effect, which may make the final position of the ring indeterminate. Due to the curved surface and necessity of the ring at the center of the sphere it can be expected that friction will have a greater effect than in the corresponding model for E^2. Also, the non-convexity of the problem results in the possibility of only getting local minima and missing the global optimum.

As mentioned by Eilon, et al. (1971), perhaps the greatest disadvantage is the fact that the method does not evaluate the cost function. The next model resolves this difficulty and essentially eliminates the above mentioned disadvantages.

III.5.2 Electronic Analogue

Hitchings (1969) recognized that most of the literature on location problems to that date was confined to problems in the planar state. He addressed the solution of problems in E^3 and developed an electronic analogue model to solve these problems.

Hitchings' model can effectively be modified to solve the problem in S^2 space using the great circle metric. One must utilize a non-conductive sphere, set up as in the mechanical analogue model. The string, however, is replaced by resistive wire in which distance is proportional to the length of the resistor. Weights are needed solely to insure that the wire conforms to a geodesic path on the sphere. The varying weights of the demand points are accommodated via changes in resistivity of material or cross-sectional area of the wire. As pointed out by Hitchings, such a model can even handle nonlinear costs by segmenting the wires into appropriate lengths and varying resistances.

Based upon a slide-wire concept, the model's circuit diagram (Hitchings 1969) is similar to a Wheatstone bridge circuit. The difference lies in the fact that the objective is to maximize current flow rather than find a null point. Note that for a fixed voltage, maximization of current flow is equivalent to minimization of overall resistance, which in turn is the analogue of the objective function for the single facility location problem on the sphere.

Attaching a pen to the ring on the sphere's surface, isocost lines may be easily plotted by moving the pen while carefully maintaining a constant current flow. In this manner contours can be plotted to reveal local minima. From this information it is simple to determine the location of the global optima (point(s) of maximum current flow).

The advantages of this model over the previous one are obvious. Relative cost function values are available via current flow values, friction has no effect, and with a little care problems for regions larger than a hemisphere can be handled. Since contours are available, the model may be useful when considering practical problems with geographic infeasibilities.

III.6 Steiner's Problem and Fagnano's Result

III.6.1 Introduction

Consider the problem of determining for a triangle in the plane the point at which the sum of distances from the point to the vertices is minimized. This is a special case of the single facility location problem where all weights are equal. It is known that if any angle of the triangle equals or exceeds 120^o, the optimum occurs at the vertex of the obtuse angle. If all angles are less than or equal to 120^o, the optimum is interior to the triangle at a point at which each side of the triangle subtends an angle of 120^o. A recent elegant proof of the former property has been given by Sokolowsky (1976). Proofs of both properties are provided by Courant and Robbins (1941), among others.

Called Steiner's Problem by many, it has a long history which is succinctly outlined by Cooper (1963). Also mentioned by Cooper is a result due to Fagnano in 1775 showing that the point for which the sum of the distances from the vertices of a quadrilateral is a minimum is given by the intersection of the diagonals. The purpose of this section is to consider extension of the above results to the problem on the sphere.

III.6.2 Steiner's Problem on the Sphere

Consider three demand points on the sphere with equal weights. Suppose an optimal solution, local or global, is known. Using the optimal point as pole, transform the points to the plane with an Azimuthal Equidistant Projection (Deetz and Adams 1948). Such a projection preserves distance and bearing to other points from the polar point.

The image of the polar point transformed from the sphere is globally optimal on the plane. Suppose it were not optimal on the plane. Since the problem is convex (Love 1967), a move could be made over an arbitrarily small distance $\delta > 0$ in some direction and yield an improvement in the objective function value. Since it can be shown that the objective function value on the plane is greater than value on the sphere for the corresponding point (see Section III.10.2) there would be an improvement by moving in the same direction and distance on the sphere. This contradicts the fact that a local or global optimum is reached on the sphere.

Since the point is global on the plane, it is known that all angle measures subtended from the triangle sides are $\geq 120^{\circ}$, with strict inequality holding only when the point is a vertex. The projection used preserves angles at the polar point so the property is retained at the inverse image--which is a local or global optimum on the sphere.

It is important to note this result is a necessary condition for global optimality on the sphere. It is suspected, but not proven, that if the points are such that no side of the spherical triangle is of length $\geq \pi/2$, then if any vertex angle is 120° or larger, its vertex is the global optimum. Otherwise, the global optimum is at the interior point at which each side of the triangle subtends an angle of 120°.

III.6.3 Fagnano's Result on the Sphere

Suppose four equally weighted points determining a convex quadrilateral (Kay 1969) are on an open hemisphere and the optimal point is $X_s{}^*$. The optimal point occurs at the intersection of the small great circle arcs.

Given the optimal point, project to the plane via the Azimuthal Equidistant Projection. Distance and bearing are preserved from point X_p on the plane corresponding to $X_s{}^*$.

On the plane the optimal is determined by intersection of the diagonals. The optimal point must coincide with X_p. For if X_p is not optimal on the plane, then improvement can be made by moving an arbitrarily small distance $\delta > 0$ in some direction from X_p to a point X'. Now, as stated in the previous section, the objective function value of X' for the Euclidean norm is strictly greater than the objective function value obtained for the great circle metric using the point corresponding to X' for the problem on the sphere.

This means there is a point on the sphere dominating the global optimum--a contradiction.

So X_p and $X_s{*}$ are optimal on the plane and sphere respectively--and are corresponding points via the transformation.

Through the properties of the projection, the angles and distances are preserved at X_p and $X_s{*}$. It follows immediately that the desired property of the global solution at X_p extends to the sphere's global solution as a necessary condition.

III.7 A Conjecture Concerning Global Optimality

In working with the problem on the sphere, and after plotting a number of example problems three dimensionally, an interesting characteristic came to light. Although not established theoretically, it is important to report since its validity would significantly impact on conclusions concerning global or local optimality.

Conjecture 3.7.1 If (1) all demand points for problem (3.2.1) are located within an octant of the unit sphere, or a disk of diameter $\leq \pi/2$ and (2) at least three of the demand points are non-collinear, then the objective function is unimodal within the region.

The conjecture incorporates two fundamental assumptions. First, that there exists a unique global minimum within the region. This is assured by requiring non-collinearity of demand points. Second, the characteristic of the objective function under this assumption is unimodal. An explicit discussion of generalized unimodality in n-dimensions is provided by Sivazlian and Stanfel (1975). A function f is said to be unimodal over a region S if there exists a path from $x \in S$ to the global optimum x^* over which f is strictly decreasing.

If this conjecture is true, then a convergent algorithm searching over the region will result in a global optimum. This follows since the dominance results of Section III.4.2 permit the search for a global to be restricted to the spherically convex hull of the demand points. The conjecture eliminates the possibility of having local minima within the region.

A typical contour of a problem with the given properties is found in Figure 3.6. It plots the objective function values for the six point unequal weight minisum problem using the Data Set D3 of Appendix B.

Concerning the disk of diameter $\pi/2$ and the octant, both regions can contain sets of demand points not contained in the other. The octant, also known as the Reuleaux

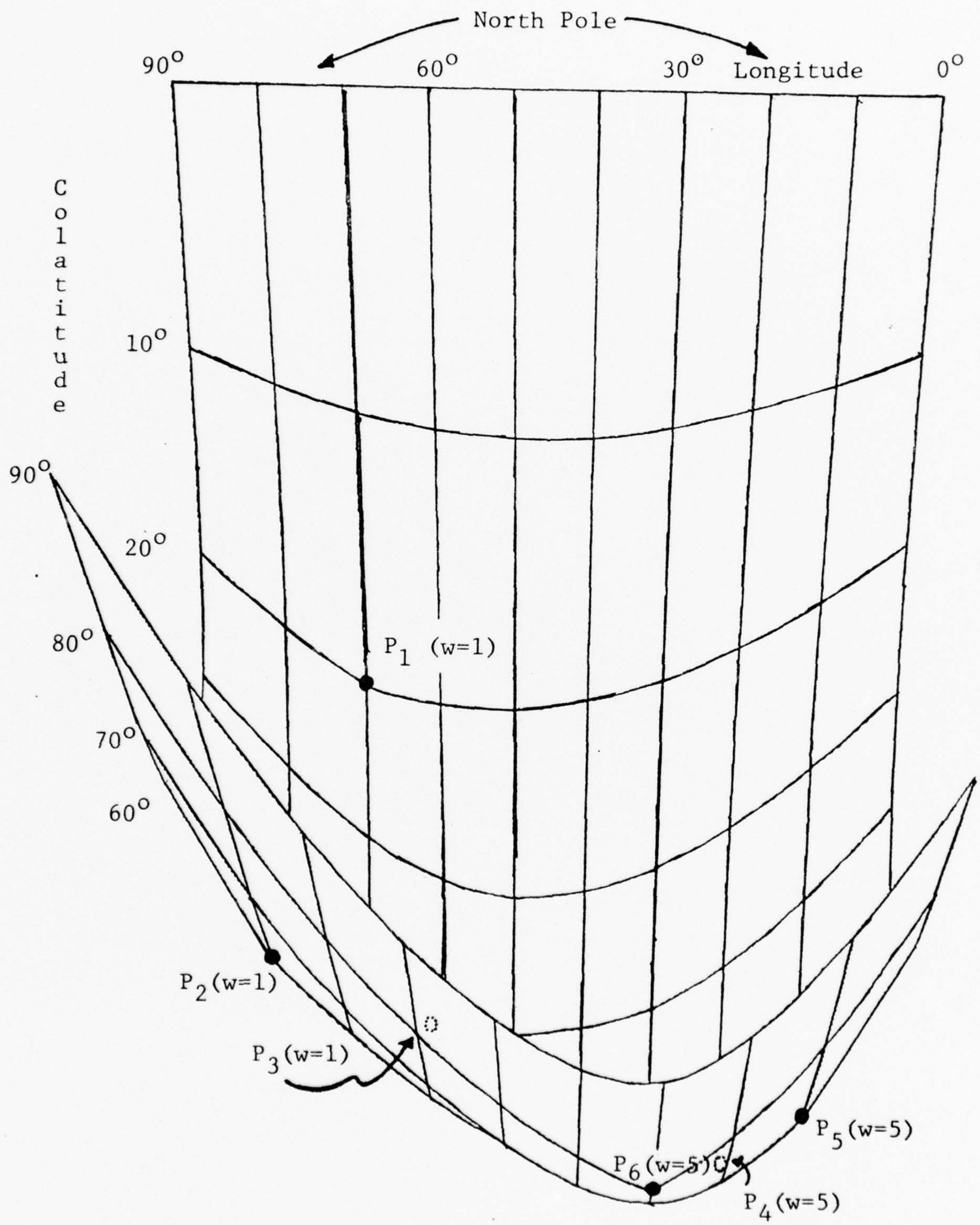


Figure 3.6. 3-D minisum functional plot; Data Set D3, Appendix B.

Triangle on the sphere, has the smallest area of all figures of equal altitude. Using techniques of Chapter IV it is possible to determine whether the demand points are located in a disk of diameter $\leq \pi/2$. It is not so easy to determine analytically whether all points are contained with an octant. However, one can always plot the points on a sphere and use an overlay.

It is suspected that the unimodality property is related to the fact that many Euclidean-like properties hold in such regions (Kay 1969). These properties hold since no distance between any demand points or between a possible location and any demand point is greater than or equal to $\pi/2$. In such regions, for example, it is known that the hypotenuse of a spherical right triangle is the longest side. This property, among others, does not hold in the general case.

III.8 An Approximate Solution

Wendell (1971), formulating the problem as (3.2.3), approximated the objective function by recognizing that Arcsin $y \approx \pi y^2/2$ for $y \in [0,1]$. Through Schwartz's inequality, the solution to the revised problem is seen to be

$$x_0^* = \begin{pmatrix} \sum_{i=1}^{M} w_i P_i^1 \\ \sum_{i=1}^{M} w_i P_i^2 \\ \sum_{i=1}^{M} w_i P_i^3 \end{pmatrix} \frac{1}{\sqrt{(\sum_{i=1}^{M} w_i P_i^1)^2 + (\sum_{i=1}^{M} w_i P_i^2)^2 + (\sum_{i=1}^{M} w_i P_i^3)^2}}$$

Wendell offers x_0^* as an approximate solution to (3.2.1).

Through simple algebraic manipulation it can be shown that if the centroid is found in E^3 and projected to the spherical surface (normalized in the case of the unit sphere), the resulting point is equivalent to x_0^*. That is, Wendell's approximate solution is the projected centroid.

As will be seen in Chapter V, it is not difficult to construct examples in which the projected centroid is far from the optimal solution, nor can one always successfully utilize it as a starting solution in any iterative technique in order to find a global optimum. It is not without value, however, as will be discussed later.

III.9 Bounds for Unconstrained Objective Function

In this section bounds are generated for the unconstrained great circle metric single facility location problem. The results are an extension of those due to Pritsker and Ghare (1970) for the Euclidean problem.

The bounds are based upon optimal solutions for the rectilinear and Euclidean norm single facility problem in E^3. As mentioned earlier, efficient procedures exist for solving these problems.

The bounds are:

$$[R^2(x^R)+R^2(y^R)+R^2(z^R)]^{\frac{1}{2}} \leq A(X^0) \leq A(X^R) \qquad (3.9.1)$$

$$E(X^E) \leq A(X^0) \leq A(X^E) \qquad (3.9.2)$$

where: $X^R = (x^R, y^R, z^R)$ is the optimum rectilinear solution,
$X^E = (x^E, y^E, z^E)$ is the optimum Euclidean solution,
$X^0 = (x^0, y^0, z^0)$ is the optimum great circle metric solution,

and R(x), E(x), and A(x) are the objective function values for the rectilinear, Euclidean and great circle metric problems, respectively.

Note that although the bounds in (3.9.2) are more "costly" to compute, they will in general be the tighter of the two. This becomes evident in the derivations which follow:

The right hand inequalities are both clearly true for the optimum of the objective to minimize A is X^0, so $A(X^0) \leq A(X)$ for all X.

In particular, $A(X^0) \leq A(X^R)$ and

$A(X^0) \leq A(X^E)$.

The fact that $A(X^0) \geq E(X^0)$ is intuitive, but can be established geometrically. Concerning the relationship between arc $\widehat{ANB}$ and chord $\overline{ASB}$ (see Figure 3.7), note that $Z = \frac{1}{2}\,\widehat{ANB}$, and $\overline{SB} = \frac{1}{2}\,\overline{ASB}$.

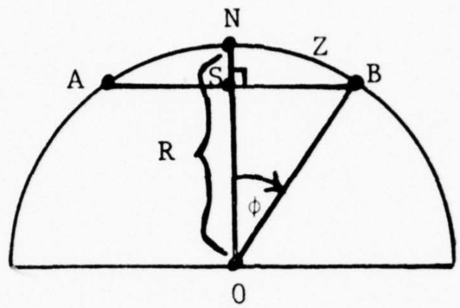


Figure 3.7. Relationship between chords and arcs.

Now, without loss of generality ϕ can be limited to the region $[0,\pi/2]$. Consider the right triangle NSB where $\measuredangle NSB = \pi/2$. Since the hypotenuse of a right triangle is the longest side, $\overline{NB} \geq \overline{SB}$. The shortest distance between any two points is a straight line so $\widehat{NB} \geq \overline{NB}$, and it follows that $Z = \widehat{NB} \geq \overline{SB}$. So $\overline{ASB} \leq \widehat{ANB}$.

Since $A(X^0)$ is a weighted sum of great circle arcs of form $\widehat{ANB}$ and $E(X^0)$ is the corresponding Euclidean sum, $A(X^0) \geq E(X^0)$.

Clearly $E(X^E) \leq E(X^0)$. Otherwise the optimality of X^E for the Euclidean norm is contradicted. This establishes (3.9.2).

For (3.9.1), application of Minkowski's inequality establishes that $E(X^E) \geq [R^2(x^R) + R^2(y^R) + R^2(z^R)]^{\frac{1}{2}}$ and the bound follows via transitivity.

Since X^R is optimum for the rectilinear problem one has that

$$[R^2(x^R)+R^2(y^R)+R^2(z^R)]^{\frac{1}{2}} \leq [R^2(x^E)+R^2(y^E)+R^2(z^E)]^{\frac{1}{2}}$$

Now,

$$[R^2(x^E)+R^2(y^E)+R^2(z^E)]^{\frac{1}{2}} = [(\sum_{i=1}^{M} w_i|x^E-a_i|)^2 + (\sum_{i=1}^{M} w_i|y^E-b_i|)^2 + (\sum_{i=1}^{M} w_i|z^E-c_i|)^2]^{\frac{1}{2}}$$

which, by Minkowski's inequality is

$$\leq \sqrt{(w_1|x^E-a_1|)^2+(w_1|y^E-b_1|)^2+(w_1|z^E-c_1|)^2} + \ldots$$

$$+\sqrt{(w_M|x^E-a_M|)^2 + (w_M|y^E-b_M|)^2+(w_M|z^E-c_M|)^2}$$

$$= \sum_{i=1}^{M} w_i \sqrt{(x^E-a_i)^2+(y^E-b_i)^2+(z^E-c_i)^2} = E(x^E,y^E,z^E)$$

$$= E(X^E)$$

III.10 A Planar Projection Algorithm (PPA)

III.10.1 Introduction

In this section the first of two algorithms for the minisum single facility location problem is developed. It capitalizes on existing solution techniques for the planar case and makes use of the fact that isometric (distance-preserving) transformations from the sphere to the plane from a single point are possible. The appropriate transformation is well-known to geographers as the Azimuthal Equidistant Projection.

As will be seen, a fundamental step in the algorithm is based upon an iterative technique for the Euclidean norm problem developed independently by Cooper (1963), Kuhn and Kuenne (1962), and Weiszfeld (1936). A modified algorithm is used; namely, the Hyperboloid Approximation Procedure (HAP) due to Eyster et al. (1973), which circumvents the difficulties of non-differentiability at demand points. In the plane, the original iterative procedure for the single facility problem is guaranteed to converge to the optimum location. A discussion of convergence properties is given by Weiszfeld (1936), Katz (1969, 1974) and Kuhn (1973). The iterative procedure has been found by Cooper (1963) and Eyster et al. (1973), among others, to solve the problem with excellent results. With this in mind, it would seem advantageous to employ the technique in solving the problem on the sphere.

The basic steps of the approach follow:

(X^S on sphere, X^P on plane)

STEP 0: Designate starting point $X_0{}^S$. Set k = 1. Set stopping criterion parameter, ε.

STEP 1: Perform Azimuthal Equidistant Projection to Euclidean plane using X^S_{k-1} as the polar point; note that the point corresponding to X^S_{k-1} is $X^P_{k-1} = (0,0)$.

STEP 2: Employ HAP algorithm using Euclidean norm to find global optimum X^P_k in plane. If $|X^P_k - X^P_{k-1}|_{\ell_2} < \varepsilon$, stop. Otherwise, go to Step 3.

STEP 3: Perform inverse transformation on X^P_k and return to sphere to get X^S_k. If $|\phi(X^S_k) - \phi(X^S_{k-1})| < \varepsilon$, where ϕ

is the objective function, stop. Otherwise, let $k = k+1$ return to STEP 1.

In the paragraphs that follow, the mathematical bases for STEP 1 and STEP 3 will be examined. STEP 0 is discussed in Section III.12. For a detailed discussion of HAP (STEP 2) refer to Eyster et al. (1973).

A specific advantage of this approach is that while solving the problem on the plane one is working with a convex problem. Via the projection one can consider problems with demand points scattered around the entire sphere without using approximations to great circle distances. Also, under the projection the number of non-differentiable points is halved since there is no problem with antipodal points on the plane.

III.10.2 Convergence Properties

As stated, it is known that the general iterative technique converges to a global optimum on the plane. Unfortunately, convergence of an iterative algorithm can only be guaranteed to a local minimum in the general problem on the sphere due to its non-convexity. Assuming a convergent algorithm is used in the plane, this guarantee can be made by verifying that each iteration of the algorithm will result in a strict improvement of the objective function value. That is, each time one projects to the plane and then returns to the sphere, any movement in the location of the servicing facility will result in a strict decrease in total cost.

Starting with X_0 on the sphere as a pole, project to the plane using the Azimuthal Equidistant Projection. Suppose X^* is the global optimum on the plane for the projected demand points. Assume it is distinct from the image of X_0, X_p. That is, assume X_p is not globally optimal on the plane. If d_i is the image of demand point P_i for each i and X_s is the inverse image of the planar optimum X^* on the sphere, it follows from the cosine inequality for elliptic geometry (Kay 1969) that (see Figure 3.8):

$$|X^*-d_i|_{\ell_2} > \rho(X_s,P_i) \quad \forall_i, \; i = 1,\ldots,M$$

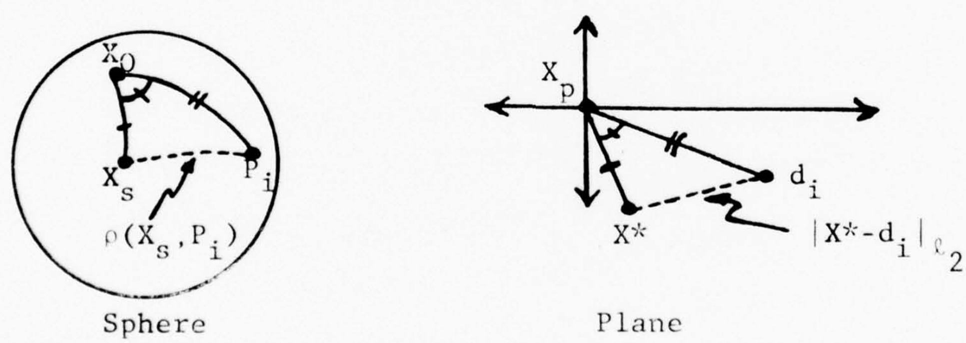


Figure 3.8. Cosine inequality for elliptic geometry.

All geodesic distances and angles from X_p are isometric to the corresponding great circle distances and angles on the sphere. All other distances are distorted on the plane to larger than the actual distance.

For the inverse image point X_s, then,

$$\sum_{i=1}^{M} w_i |X^* - d_i|_{\ell_2} > \sum_{i=1}^{M} w_i \rho(X_s, P_i)$$

Also, due to properties of the projection, and the fact that X* is global on the plane it follows that

$$\sum_{i=1}^{M} w_i \rho(X_0, P_i) = \sum_{i=1}^{M} w_i |X_p - d_i|_{\ell_2} \geq \sum_{i=1}^{M} w_i |X^* - d_i|_{\ell_2}$$

Thus $\sum_{i=1}^{M} w_i \rho(X_0, P_i) > \sum_{i=1}^{M} \rho(X_s, P_i)$ and strict improvement is guaranteed.

The question arises as to when one can expect the location of the servicing facility to move during the search for a global optimum on the plane. For, if no movement is made from the polar point X_p, there will be no change in the objective function value. With this in mind, it is shown that the characteristics of the first partial derivatives on the sphere and on the plane are similar. That is, if the partial derivatives are non-zero in the sphere problem, they will likewise be non-zero in the planar problem. By employing a modified gradient approach in the spirit of Kuhn (1973), there is no difficulty in assuming differentiability even in the situation when the polar point is a demand point or its antipodal point.

Now, the convergent algorithm in the plane will result in movement from a non-stationary point to a stationary point. So if the foregoing is established one is guaranteed of achieving a strict improvement in any iteration which projects from a non-stationary point on the sphere.

Theorem 3.10.1 Given an Azimuthal Equidistant Projection from a point X_0 on the sphere, let X_p be the corresponding point on the plane via the projection. Suppose the first partial derivatives of the objective function on the sphere exist and are non-zero at X_0. Then the first partial derivatives of the planar objective function are likewise non-zero at X_p.

Proof: The approach is via contradiction. Without loss of generality, the coordinates on the sphere can be translated so that $X_s = (\phi_x, \lambda_x) = (0,0)$ in colatitude, longitude. Under the projection, of course, $X_p = (X_x, Y_x) = (0,0)$.

Suppose the first partial derivatives at X_p vanish in the plane. The proof will be established by showing that the first partial derivatives vanish for the corresponding point $X_s = (0,0)$ on the sphere.

Since the Euclidean metric is used in the plane we have

$$\left.\frac{\partial f(X_x, Y_x)}{\partial X_x}\right|_{(0,0)} = \sum_{i=1}^{M} \frac{w_i(X_x - a_i)}{[(X_x-a_i)^2+(Y_x-b_i)^2]^{\frac{1}{2}}} = \sum_{i=1}^{M} \frac{-w_i a_i}{(a_i^2+b_i^2)^{\frac{1}{2}}}$$

$$= 0$$

$$\left.\frac{\partial f(X_x, Y_x)}{\partial Y_x}\right|_{(0,0)} = \sum_{i=1}^{M} \frac{w_i(Y_x - b_i)}{[(X_x-a_i)^2+(Y_x-b_i)^2]^{\frac{1}{2}}} = \sum_{i=1}^{M} \frac{-w_i b_i}{(a_i^2+b_i^2)^{\frac{1}{2}}}$$

$$= 0$$

Now, under the projection it is seen that for a point $P_i = (\phi_i, \lambda_i)$ on the sphere, the corresponding point in Cartesian coordinates is (a_i, b_i) where $a_i = \phi_i \cos\lambda_i$, $b_i = \phi_i \sin\lambda_i$. This, of course, is contingent upon the projection being from $X_s = (0,0)$.

By substitution:

$$\sum_{i=1}^{M} \frac{w_i a_i}{(a_i^2+b_i^2)^{\frac{1}{2}}} = \sum_{i=1}^{M} \frac{w_i \phi_i \cos\lambda_i}{[(\phi_i\cos\lambda_i)^2+(\phi_i\sin\lambda_i)^2]^{\frac{1}{2}}} = \sum_{i=1}^{M} w_i\cos\lambda_i$$

$$= 0$$

Similarly, $\sum_{i=1}^{M} w_i \sin\lambda_i = 0$

Using the form of the objective function as (3.2.7), the following necessary conditions arise:

$$\frac{\partial\psi}{\partial\phi} = \sum_{i=1}^{M} \frac{-w_i(-\cos\phi_i\sin\phi+\sin\phi_i\cos\phi\cos(\lambda-\lambda_i))}{\sqrt{1-(\cos\phi_i\cos\phi+\sin\phi_i\sin\phi\cos(\lambda-\lambda_i))^2}}$$

$$\frac{\partial\psi}{\partial\lambda} = \sum_{i=1}^{M} \frac{-w_i(\sin\phi_i\sin\phi[\sin\lambda\cos\lambda_i-\cos\lambda\sin\lambda_i]}{\sqrt{1-(\cos\phi_i\cos\phi+\sin\phi_i\sin\phi\cos(\lambda-\lambda_i))^2}}$$

Looking specifically at when $(\phi,\lambda) = (\phi_x, \lambda_x) = (0,0)$ it is seen that:

$$\left.\frac{\partial\psi}{\partial\phi}\right|_{(0,0)} = \sum_{i=1}^{M} \frac{-w_i(\cos\phi_i\sin 0+\sin\phi_i\cos 0\cos(-\lambda_i))}{\sqrt{1-(\cos\phi_i\cos 0+\sin\phi_i\sin 0\cos(-\lambda_i))^2}}$$

$$= \sum_{i=1}^{M} \frac{-w_i\sin\phi_i\cos\lambda_i}{\sqrt{1-\cos^2\phi_i}} = \sum_{i=1}^{M} - w_i\cos\lambda_i = 0$$

Also,

$$\left.\frac{\partial\psi}{\partial\lambda}\right|_{(0,0)} = \sum_{i=1}^{M} \frac{-w_i(\sin\phi_i \sin 0[\sin 0\cos\lambda_i - \cos 0\sin\lambda_i])}{\sqrt{1-\cos^2\phi_i}} = 0$$

This completes the proof.

The question of whether the algorithm converges when used with a convergent algorithm in the plane is not really appropriate in this case since the planar search is done with an efficient algorithm that uses a heuristic as a stopping criterion.

III.10.3 The Azimuthal Equidistant Projection

In this section the mathematical formulations used for projecting to the planar surface are developed. This is essentially STEP 1 of the three major steps outlined in Section III.10.1.

Letting (ϕ,λ) represent the spherical coordinates in colatitude $(0\leq\phi\leq\pi)$ and longitude $(-\pi<\lambda\leq\pi)$, the projection from the North Pole to the plane is trivial. Using polar coordinates, (r,θ), the mapping transformation is simply:

$$r = \phi$$

$$\theta = \lambda$$

where 0^{o} longitude, the Greenwich Meridian, transforms to the positive X axis in a Cartesian system, and the South Pole is a singular point arbitrarily placed at $(\pi,0)$. The Cartesian representation (x,y) is then:

$$x = \phi\cos\lambda$$

$$y = \phi\sin\lambda$$

As mentioned earlier, this particular projection preserves actual distances and bearing from the North Pole. Error increases as one determines distances between points non-collinear with a ray from the origin and some distance from the pole. Recognizing this, the algorithm incorporates a bound which stops the planar search whenever the movement exceeds a fixed distance from the pole.

III.10.4 Transformation of Poles

In the Planar Projection Algorithm (PPA) one is unfortunately hardly ever projecting from the North Pole. It is desired to project from points all over the sphere, namely, the best point achieved in the previous iteration. This requires a method to transform the spherical coordinate system so that the "North Pole" is at the desired pole for projection. The transformation to the plane, then, is trivial.

Using a concept familiar to navigators, it is possible to transform the coordinate system without ambiguity from one pole to another. The concept to be utilized is the haversine, where

$$\text{hav}\theta = \tfrac{1}{2}(1-\cos\theta) = \sin^2(\theta/2)$$

Transferring from the North Pole to new pole (θ_0,λ_0), the new coordinates (Z_i,α_i) for each demand point $P_i = (\phi_i,\lambda_i)$

are given by Maling (1973):

$$\text{hav}Z_i = \text{hav}|\phi_i-\phi_0| + \sin\phi_i\sin\phi_0\text{hav}|\lambda_0-\lambda|$$

$$\text{hav}\alpha_i = (\text{hav}\phi_i-\text{hav}|\phi_0-Z_i|)\csc\phi_0\csc Z_i$$

Using inverse trigonometric functions, one can solve for (Z_i,α_i), the bearing and distance coordinates from the new pole to all demand points P_i, $i=1,\ldots,M$.

The end result, after projection to the plane, is a planar representation of the sphere with true distances from the new pole to all points P_i and only "small" error in distances from other locations to the P_i in regions near the pole.

Solving for Z_i:

$$\text{hav}Z_i = \text{hav}|\phi_i-\phi_0| + \sin\phi_i\sin\phi_0\text{hav}|\lambda_0-\lambda_i|$$

$$\tfrac{1}{2}(1-\cos Z_i) = \sin^2\left(\frac{\phi_i-\phi_0}{2}\right) + \sin\phi_i\sin\phi_0\sin^2\left(\frac{\lambda_0-\lambda_i}{2}\right)$$

Thus:

$$Z_i = \text{Arccos}\{1-2(\sin^2\left(\frac{\phi_i-\phi_0}{2}\right) + \sin\phi_i\sin\phi_0\sin^2\left(\frac{\lambda_0-\lambda_i}{2}\right))\} \quad (3.10.1)$$

Solving for α_i:

$$\text{hav}\alpha_i = (\text{hav}\phi_i - \text{hav}|\phi_0-Z_i|)\csc\phi_0\csc Z_i$$

$$\tfrac{1}{2}(1-\cos\alpha_i) = [\sin^2\left(\frac{\phi_i}{2}\right) - \sin^2\left(\frac{\phi_0-Z_i}{2}\right)]\csc\phi_0\csc Z_i$$

Thus:

$$\alpha_i = \text{Arccos}\{1-2[\sin^2\left(\frac{\phi_i}{2}\right) - \sin^2\left(\frac{\phi_0-Z_i}{2}\right)]\csc\phi_0\csc Z_i\} \quad (3.10.2)$$

Although α_i is undefined when either ϕ_0 or Z_i are 0 or π this is easily resolved by arbitrarily defining α_i to be 0 since one is at the North or South Pole. Z_i is defined as either 0 or π, whichever is correct.

Observing that the range of α_i is $-\pi < \alpha_i \leq \pi$, it is obvious that there will be ambiguity as to the proper sign due to the range of $[0,\pi]$ for the Arccos function. Now, the azimuth α_i is in reference to the shortest great circle arc from the new pole to the old pole, which is the new zero meridian. In order to determine the proper sign it is necessary to check the sign of $(\lambda_0-\lambda)$. The cases in Table 3.1 apply.

	$\lvert\lambda_0-\lambda\rvert \leq \pi$	$\lvert\lambda_0-\lambda\rvert > \pi$
$\lambda_0-\lambda \geq 0$	$\alpha_i = \alpha_i$	$\alpha_i = -\alpha_i$
$\lambda_0-\lambda < 0$	$\alpha_i = -\alpha_i$	$\alpha_i = \alpha_i$

Table 3.1. Sign of Azimuth.

An example of the case where $\lvert\lambda_0-\lambda\rvert \leq \pi$ and $\lambda_0-\lambda \geq 0$ is seen in Figure 3.9.

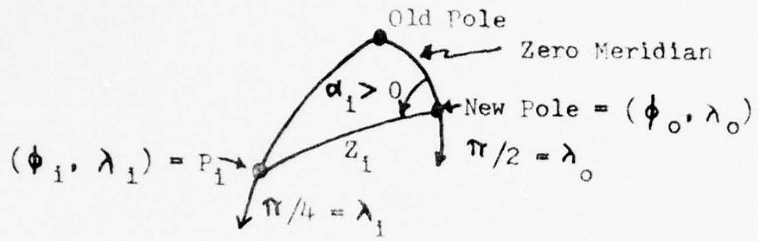


Figure 3.9. Determination of Azimuth (longitude) sign.

III.10.5 Returning to the Sphere

Consider now STEP 3 of the three major steps in Section III.10.1. Prior to beginning the next iteration and after optimization on the plane it is necessary to return to the sphere. This is done by converting the globally optimal point (X^*,Y^*) to spherical coordinates using

$$\phi = [(X^*)^2+(Y^*)^2]^{\frac{1}{2}} \qquad (3.10.3)$$
$$\lambda = \mathrm{Arctan}(Y^*/X^*)$$

Ambiguity regarding the sign of λ is easily handled. Once on the sphere it is simple to transform the pole back to the North Pole, using the procedure of STEP 1. In this way the point (X^*,Y^*) is expressed in spherical coordinates with the North Pole as pole.

III.10.6 The Algorithmic Procedure

Utilizing the results and mathematical development of the previous section, the following algorithm is obtained.

Terminology and notation used in any iteration, ITER, are:

BOUND = upper bound for movement of search on plane before returning to sphere

EP1 = stopping criterion for movement on the plane in one iteration

EP2 = stopping criterion for change in objective function value on the sphere using (EP2)*F, where F is the current objective function value.

K = (0,1). K = 0 if starting solution is to be projected centroid. K = 1 otherwise.

N1 = bound on major iterations

M = number of demand points

W_j = weight of demand point j

ϕ_j = colatitude of demand point j in radians

λ_j = longitude of demand point j in radians

1. Input parameters BOUND, EP1, EP2, K, N1, M and demand points (ϕ_j, λ_j) with weights W_j for $j = 1,\ldots,M$.

2. If K = 0 go to step 3. Otherwise, go to step 4.

3. Calculate the projected centroid $(x\phi_0, x\lambda_0)$ for starting solution. Go to step 5.

4. Input $(x\phi_0, x\lambda_0)$.

5. Using (3.2.6) calculate initial objective function value on the sphere letting $(x\phi_0, x\lambda_0)$ be the servicing facility. Call it OBSPH.

6. Set ITER = 1, $x\phi = x\phi_0$, and $x\lambda = x\lambda_0$.

7. If ITER $\geq$ N1, go to step 18. Otherwise, go

to step 8.

8. Transform pole to $(x\phi, x\lambda)$. Designate new coordinates of demand points as $(\phi T_j, \lambda T_j)$ for $j = 1,...,M$. Coordinates of North Pole are $(x\phi_{NP}, x\lambda_{NP})$.

9. Set F_B = OBSPH.

10. Transform all demand points $(\phi T_j, \lambda T_j)$ from spherical coordinates to Cartesian coordinates via the Azimuthal Equidistant Projection. Designate them as (A_j, B_j).

11. Using the Hyperboloid Approximation Procedure (HAP), or any other Euclidean norm single facility algorithm, solve the problem in E^2 for the global minimum (X^*, Y^*).

12. Check to see if the first stopping criterion is satisfied. If $[(X^*)^2+(Y^*)^2]^{\frac{1}{2}} <$ EP1 then do steps 13 through 15 and go to step 18. Otherwise, go to step 13.

13. Transform (X^*, Y^*) from Cartesian to spherical coordinates, thus returning to the sphere with $(x\phi, x\lambda)$ as pole to get $(x\phi^*, x\lambda^*)$.

14. Transform the pole at $(x\phi, x\lambda)$ back to the North Pole $(x\phi_{NP}, x\lambda_{NP})$ to get $(x\phi^*, x\lambda^*)$ in terms of the North Pole. Call the new point $(x\phi, x\lambda)$.

15. Using (3.2.6) evaluate the objective function OBSPH on the sphere, letting $(x\phi, x\lambda)$ be the servicing facility. Go to step 16 (unless stopping criterion in step 12 is satisfied).

16. Set F_A = OBSPH.

17. Check to see if the second stopping criterion is satisfied. If $|F_A - F_B| < (EP2)*F_A$, go to step 18. Otherwise, go to step 7.

18. Stop.

III.11 A Cyclic Meridian Search Algorithm (CMS)

III.11.1 Introduction

Recognizing that an appreciable amount of time is required for calculation of the gradient in the planar search and in transforming back and forth from the sphere, one can reasonably consider the possibility of both searching entirely on the spherical surface and avoiding the use of derivatives. Another factor to encourage such an approach is that due to the way a gradient search functions, it will proceed downhill to a local optimum once it is within its "region of attractiveness." A derivative-free approach may be able to avoid local optima.

Two questions come immediately to mind. That is, how should one choose a search direction, and how should the line search be made. Two simple procedures were used.

III.11.2 The Search Direction

Concerning search directions, the developed algorithm always searches, in each cycle, in a direction orthogonal to a meridian along a great circle track and then along a meridian. This cycle is repeated until an initial stopping criterion is satisfied. At that juncture random search directions are tried in order to further improve the objective function value. This procedure helps to avoid getting stranded on ridges and in valleys--as can happen in a cyclic coordinate (one at a time) search on the plane.

The occurrence of such ridges and valleys on the sphere can best be demonstrated via example. Consider the four point single facility location problem using data set D4 of Appendix B, depicted in Figure 3.10. The global optimum in degrees of latitude and longitude is at approximately (33,57), yet a cyclic search that does not incorporate random search directions could get stuck at the point (0,20) even though it is clearly not a local minimum. This would happen if the optimization started with 20° longitude as a fixed meridian and a line search along it converged to (0,20). At this point an orthogonal search would be along the great circle arc which is identical to the Equator, or 0° latitude. Such a search would not decrease the value of the objective function so the algorithm would terminate at (0,20).

The cyclic search process can be quite slow around ridges and valleys due to a tendency of the search to zig-zag along them. This is discussed with regard to this four point problem in Chapter V. Overall, however, the algorithm seemed to be quite efficient when compared to the planar projection method.

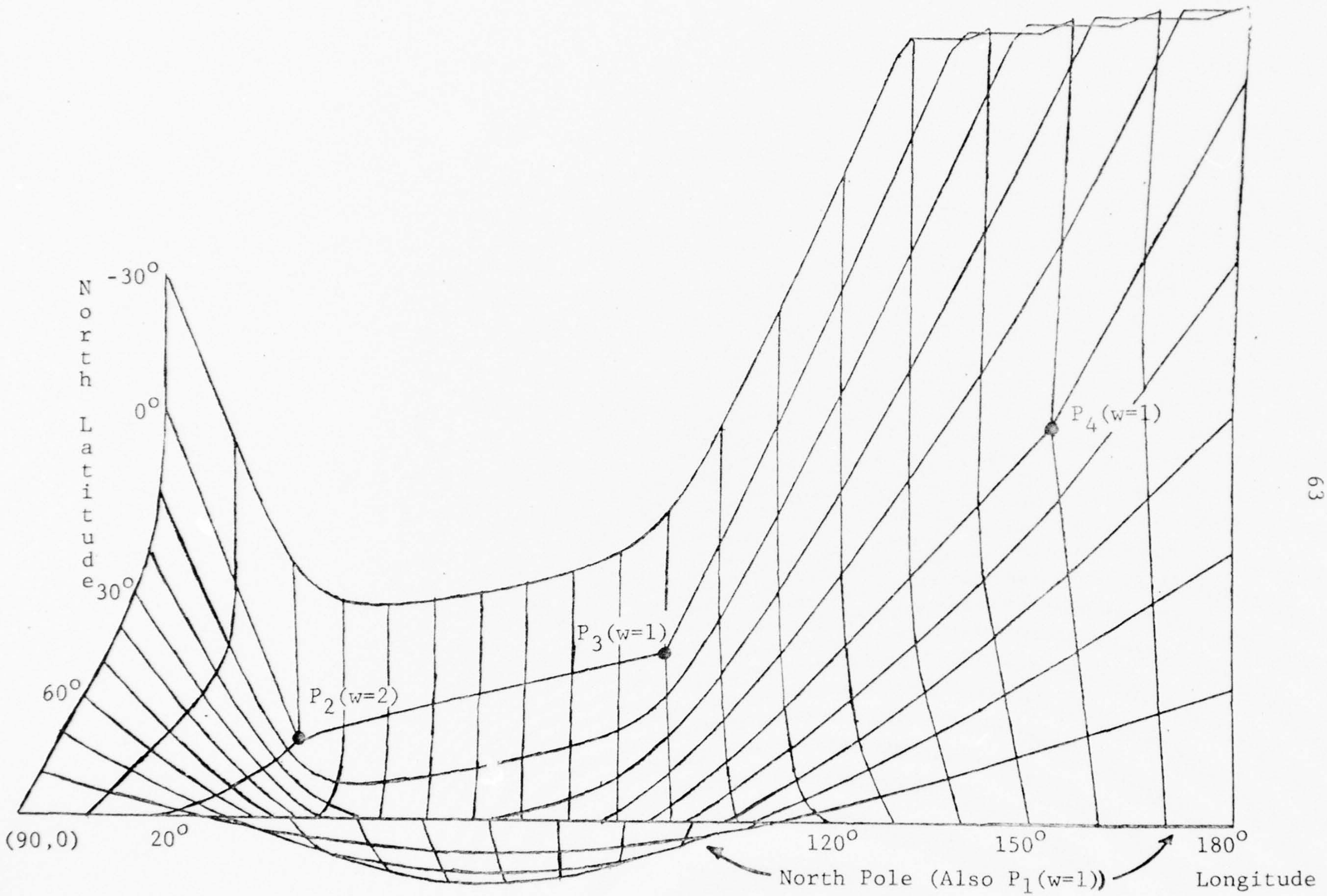


Figure 3.10. 3-D minisum function plot; Data Set D4, Appendix B.

It is easy to visualize that although search directions are orthogonal within each cycle, the actual direction of search on the sphere can be in any direction--except in a pathological case when the line search in a fixed direction moves an exact multiple of $\pi/2$ at every iteration. This property in itself would appear to make the technique more powerful than the counterpart cyclic coordinate descent algorithm on the plane (Zangwill 1974).

III.11.3 The Line Search

Regarding the line search, it must be recognized that although the search along a geodesic (meridian) in any direction from a pole is bounded by π, the objective function is not in general unimodal over the search region. As a result, the standard efficient line search techniques (Fibonnaci, Golden Section) will not guarantee a global optimum. Also, since local optima exist in the general problem, one should give thought to techniques of escaping them when at all possible. It turns out that an adaptation of a simple search technique due to Bazaraa (1975) will perform a good line search and in many cases avoid local optima, especially when the local optima's region of attraction are relatively small and the objective function's surface is not too flat.

The algorithmic steps of the line search will be described in the next section. A general outline follows:

Given the objective function where $f(\phi^*,\lambda) = Z^*$ where λ is a fixed meridian--along which the line search is being made--check to see if $f(\phi^*+\varepsilon,\lambda)$ or $f(\phi^*-\varepsilon,\lambda)$ results in a decrease in the objective function value for arbitrarily small $\varepsilon > 0$. If neither decrease the objective function value the search is terminated. Suppose it is determined that $f(\phi^*+\varepsilon,\lambda)$ is the appropriate search direction. The intent of the search then is to increase the colatitude ϕ as much as possible for fixed λ, all the while decreasing the objective function value. Choosing an initial step size $\Delta > 0$, and an initial acceleration factor $S = 1$, evaluate $f(\phi^*+S\Delta,\lambda)$. If $f(\phi^*+S\Delta,\lambda) < Z^*$, then Z^* is replaced by $f(\phi^*+S\Delta,\lambda)$, ϕ^* is replaced by $\phi^*+S\Delta$, S is replaced by αS $(\alpha>1)$, and the process is repeated until the first failure is encountered. If only $f(\phi^*-\varepsilon,\lambda)$ had decreased the objective function value, then S would be replaced by -1 and the colatitude ϕ would be decreased as much as possible.

When the first failure is encountered, the step size Δ is decreased and the process continued. This reduction in step size takes place after each failure until the minimum step size results in a failure. In this case, an approximate minimum (possible local) of f is ϕ^*.

A prominent feature of this approach is the fact that the initial step size Δ is a parameter that can be controlled so that in many cases it is possible to step out of the range of a local optimum. A graphical picture of this is in Figure 3.11. Note that the first evaluation gives $f(\phi^*+\Delta) < f(\phi^*)$ and the local minimum is escaped. It must be emphasized that selection of Δ has a pronounced effect on success in escaping local minima, and little is known a priori as to appropriate choices.

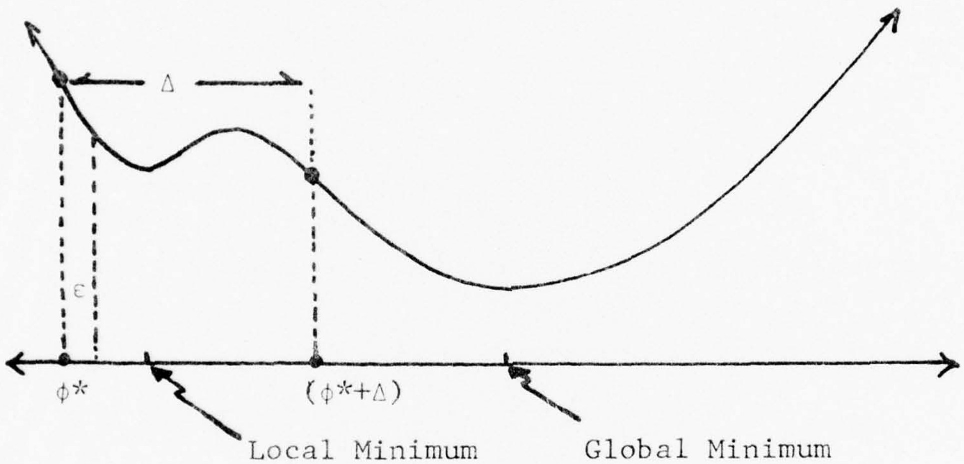


Figure 3.11. Escaping a local minimum.

III.11.4 The Algorithmic Procedure

Armed with the foregoing preliminary comments, the Cyclic Meridian Search (CMS) Algorithm is now described. The following notation and terminology is used:

S = initial acceleration factor (=1)

α = acceleration factor for step size (>1)

N = switching factor for reduction of basic step size

ϵ = small step size used to find direction ($\pm$) of search on fixed meridian

$\Delta 1$ = initial step size

ΔMIN = minimum step size

ΔINC = incremental step size

K = (0,1), K=0 if projected centroid is to be starting solution; K=1 otherwise.

M = number of demand facilities

W_j = non-negative weight associated with demand point j

ϕ_j = colatitude of demand point j in radians

λ_j = longitude of demand point j in radians

ITBOUN = bound on number of cyclic searches

EP1 = stopping criterion for search movement

EP2 = stopping criterion for improvement in objective function using EP2*(F) where F is the current objective function value

LRANBO = number of random search directions employed once either stopping criterion is satisfied

1. Input parameters α, M, $\Delta 1$, ΔMIN, ΔINC, K, ϵ, ITBOUN, EP1, EP2, LRANBO, and demand points (ϕ_j, λ_j) with weights W_j for $j = 1,\ldots,M$.

2. If K = 0, go to step 3. Otherwise go to step 4.

3. Calculate the projected centroid $(x\phi_0, x\lambda_0)$ for starting point. Go to step 5.

4. Input starting point $(x\phi_0, x\lambda_0)$.

5. Set $\phi^* = x\phi_0$, $\lambda^* = x\lambda_0$.

6. Calculate initial objective function value on the sphere using (3.2.6), letting (ϕ^*,λ^*) be the servicing facility. Call it F_{IN}.

7. Fix the meridian at λ^*, and perform a line search from ϕ^* to get $(x\phi,x\lambda)$ and F_s, the objective function value (an improvement over F_{IN}).

8. Set $F_{IN} = F_s$.

9. Set ITER = 0.

10. Set ITER = ITER + 1, LRAN = 0.

11. Set up a search from the current pole along the meridian $\pm\,\pi/2$. Set $(\phi^*,\lambda^*) = (0,\pi/2)$.

12. See if the bound on iterations has been exceeded. If ITER is less than ITBOUN, go to step 13. Otherwise go to step 33.

13. Transform the pole to $(x\phi,x\lambda)$ using procedures in Section III.10.4. Store North Pole as (ϕ_{NP},λ_{NP}). (ϕ_{NP},λ_{NP}) is required in step 28.

14. Fixing the meridian at λ^*, perform a line search from ϕ^* to get $(x\phi,x\lambda)$ and F_N, the corresponding objective function value.

15. Check to see whether a random search direction has been used within the iteration. If LRAN is greater than zero, go to step 18. Otherwise go to step 16.

16. Insure that two orthogonal searches are performed before checking stopping criteria. The two searches

complete a cycle in the two dimensional spherical space. In each cycle the first search is performed along a great circle arc orthogonal to a meridian when the North Pole is the polar point, and the second search is along a meridian. The pattern of such a search is depicted in Figure 3.12. Observe that as the region of search gets smaller, the search closely approximates a cyclic coordinate search in E^2.

If $\lambda^* = \pi/2$ go to step 17. Otherwise go to step 21.

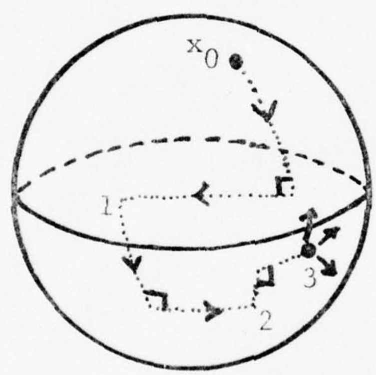


Figure 3.12. Sample search pattern for cyclic meridian search.

17. Set up values for use in testing stopping criteria after completion of second half of cycle. Set $F_1 = F_s - F_N$, $x\phi_1 = x\phi$ and $F_s = F_N$. Go to step 28.

18. If a random search direction has been used within the iteration, a stopping criterion is tested to see if it is still satisfied. If $|F_s - F_N| \leq EP2(F_N)$ go to step 22. Otherwise go to step 19.

19. Set $F_s = F_N$.

20. The second stopping criterion is tested after use of a random search direction to see if it is still satisfied. If $|x\phi| \leq$ EP1 go to step 25. Otherwise go to step 28, and begin a new iteration.

21. This step is reached after one cycle of the line search is completed--namely two orthogonal searches. A check is made to see if a stopping criterion is satisfied. If $[F_1{}^2+(F_s-F_n)^2]^{\frac{1}{2}} \leq EP2(F_N)$ go to step 22. Otherwise go to step 23.

22. Set $F_s = F_N$ and go to step 25.

23. Set $F_s = F_N$.

24. The second stopping criterion is checked after a cycle of the line search is completed. If $(x\phi_1{}^2+x\phi^2)^{\frac{1}{2}} \leq$ EP1 go to step 25. Otherwise go to step 28 and begin a new iteration.

25. This step is reached whenever any of the stopping criteria are satisfied. It enables the algorithm to make a search in a random direction. Set LRAN = LRAN + 1.

26. Check to see if the upper bound on random search directions in any iteration has been exceeded. If LRAN is greater than LRANBO go to step 28. Otherwise go to step 27.

27. Generate a random number RAN, $-1 \leq \text{RAN} \leq 1$, set $\phi^* = 0$ and let the fixed meridian (line of search) be $\lambda^* = \text{RAN}(\pi)$. Go to step 13.

28. Convert $(x\phi, x\lambda)$ to coordinates with the

North Pole as pole using (ϕ_{NP},λ_{NP}) from step 13 and procedures of Section III.10A for transforming poles. This is done in preparation for performing the next iteration or cycle.

29. Evaluate the objective function value on the sphere using (3.2.6) and letting $(x\phi, x\lambda)$ be the servicing facility.

30. If LRAN is greater than LRANBO, go to step 33. Otherwise go to step 31.

31. Determine whether the line search cycle has been completed. If $\lambda^* = \pi/2$, go to step 32. Otherwise go to step 10.

32. Begin the second half of the line search cycle. Set $(\phi^*,\lambda^*) = (0,0)$ and go to step 13.

33. Stop.

The line search procedure, being the heart of the algorithm, is now described as a separate entity.

1. Input parameters M, $\Delta 1$, ΔMIN, ΔINC, α, N, (ϕ^*,λ^*), ε and demand points (ϕ_j,λ_j) with weights W_j. Let $F_{IN} = f(\phi^*,\lambda^*)$.

2. Set ICOUNT = 0. This counter is used in determining whether the appropriate direction of search on the fixed meridian has been determined.

3. Set R = 1. R is a parameter which fixes S at a proper value, ± 1, after each change in step size Δ.

4. Set $\Delta = \Delta 1$.

5. Set S = R.

6. Set $K_0 = 0$. K_0 counts the number of accelerations for a given step size.

7. $K_0 = K_0 + 1$.

8. Check to see how many steps have been made for a particular step size Δ. This permits one to restrict the search to a maximum size interval. For example, if $\Delta 1 = .1$ radians, $S = 1$, $\alpha = 2$ and $N = 5$, the largest move along a line due to acceleration would be $2^5(.1) = 3.2$, which is just greater than π. If K_0 is greater than N, go to step 36. Otherwise go to step 9.

9. If ICOUNT = 0, go to step 27. Otherwise go to step 10.

10. Set ICOUNT = ICOUNT + 1.

11. Set $x\phi_N = \phi^*+S(\Delta)$ and $x\lambda = \lambda^*$.

12. Due to the bound on colatitude, it is necessary to insure that $x\phi_N$ is never less than zero or greater than π. If $x\phi_N$ is less than or equal to π and greater than or equal to zero, go to step 28. Otherwise, go to step 13.

Steps 13 through 22 re-orient the direction of search within the bounds of colatitude and along the proper meridian when necessary.

13. If $x\phi_N$ is less than or equal to 0, go to step 18. Otherwise go to step 14.

14. Set $x\phi_N = 2\pi - x\phi_N$.

15. If $\lambda^* + \pi$ is greater than π, go to step 17. Otherwise go to step 16.

16. Set $x\lambda = \lambda^* + \pi$ and go to step 22.

17. Set $x\lambda = \lambda^* - \pi$ and go to step 22.

18. Set $x\phi_N = -x\phi_N$.

19. If $\lambda^* + \pi$ is greater than π, go to step 21. Otherwise go to step 20.

20. Set $x\lambda = \lambda^* + \pi$ and go to step 22.

21. Set $x\lambda = \lambda^* - \lambda$.

22. Set $R = -R$ and $S = -S$.

23. Evaluate the objective function at $(x\phi_N, x\lambda)$ using (3.2.6). Call it F.

24. If ICOUNT = 0, go to step 31. This means that the appropriate direction for search along $x\lambda$ has not yet been determined. Otherwise, go to step 25.

25. Check to see if a decrease in the objective function value has been obtained. If not, then the step size must be reduced. If $(F_s - F)$ is less than zero go to step 26. Otherwise go to step 33.

26. Set $R = -R$ and $S = -S$, and go to step 37.

27. Find the appropriate direction ($\pm$) of search from ϕ^* by testing the effect on the objective function due to a small movement from ϕ^* to $\phi^* + \varepsilon$ along λ^*. Set $x\phi_N = \phi^* + \varepsilon$ and $x\lambda = \lambda^*$. Go to step 12.

28. Evaluate the objective function at $(x\phi_N, x\lambda)$ using (3.2.6). Call it F.

29. If ICOUNT = 0, go to step 31. Otherwise go to step 30.

30. Check the stopping criterion to see whether a decrease in the objective function has been obtained. If $(F_s - F)$ is less than zero, go to step 36. Otherwise go to step 33.

31. This step is entered only when determining the initial direction of search. If $(F - F_{IN})$ is less than zero, go to step 33. Otherwise go to step 32.

32. Set R = -R, S = -S, ICOUNT = ICOUNT + 1, $F_s = F_{IN}$ and go to step 11.

33. Set ICOUNT = ICOUNT + 1, $F_s = F$, $\phi^* = x\phi_N$, $\lambda^* = x\lambda$.

34. Check to see if the proper direction (±) has just been found. If ICOUNT = 1, go to step 10. Otherwise go to step 35.

35. Set $S = \alpha(S)$ and go to step 7.

36. Set $\lambda^* = x\lambda$.

37. Decrease step size Δ by ΔINC.

38. If Δ is less than ΔMIN, go to step 39. Otherwise go to step 5.

39. Set $x\phi = \phi^*$, $x\lambda = \lambda^*$.

40. Stop.

III.12 <u>Starting Solutions</u>

A fundamental consideration in any iterative technique is what to use as a starting solution. In this section the characteristics of a number of approaches will be discussed.

III.12.1 Projected Centroid

In testing the two algorithms, it was found that the projected centroid provided a "good" starting solution. Often it was in the vicinity of the global optimum, and in many cases the global optimum was achieved when this point started the iterative procedure. Finding the centroid in E^3 for the Euclidean norm and projecting it to the sphere's surface along a vector emanating from the sphere's center is easy to do analytically. As discussed in Section III.8, this point was developed as an approximate solution by Wendell. Intuitively it would seem to be a good starting solution for a problem with demand points all over the sphere since the projected centroid would at least locate one in a hemisphere containing a majority of the weighted demand.

III.12.2 Projected Optimum for Euclidean Norm in E^3

Although comparatively more expensive to determine, this point could conceivably be better than the projected centroid in certain cases. However, it unfortunately requires an iterative technique to find it.

III.12.3 Random Start

One could generate random starting points on the spherical surface. This is the basis for the most practical way to help ascertain whether global or local optima have been achieved, other than by plotting contours. In the special case when either the centroid or Euclidean distance solution in E^3 is at the center of the sphere, this procedure would be an alternative.

III.12.4 Demand Points and their Antipodal Points

Starting the iterative procedure at a demand point or its antipodal point can be informative. Since at most only a local optimum can be expected in an iterative technique, a number of different starting points will likely be tried. Use of a demand point will reveal quickly whether it is a local optimum or not. If it is, future starting points should be selected at some distance from it in order to try to avoid converging to it again. The antipodal point certainly fits this criterion.

III.13 Avoiding Local Minima

It is well known that a guarantee of an obtainable global optimum occurs only when the search region is convex and when the objective function is unimodal in the appropriate form, i.e., convex for a minimization problem. In working with a non-convex programming problem, as in this case, one must be concerned with methods of avoiding local optima. Reklaitis and Phillips (1975) state that all known techniques except those employing statistical sampling techniques will generate only local minima. They reference work by Clough (1969) and Liau (1973) concerning the statistical approach.

A standard approach when using iterative (numerical) techniques is to accept the fact that only local minima can be obtained and solve the problem a number of times using random starting points. One could then state with some degree of confidence that the global optimum has been achieved--the best solution obtained thus far. However, there is usually no way of knowing how many local minima exist and the possibility of a global optimum lying elsewhere cannot be disregarded.

Intuitively it is seen that many starting points will result in the same local optimum if they are within the "hollow" of that local minimum. Recognition of this leads to another class of approaches. One could use a solution for one problem and employ various techniques for "jumping out" of the range of the current local minimum. Hesse (1973) developed a heuristic procedure which fits in this class. A penalty function based upon an added spherical constraint is used in order to get away from a local minimum. In this way an attempt is made to find better local optima.

A third class of methods attempts to gain information about the entire search region (in the case at hand, either the spherically convex hull or the entire sphere). Attention would gradually be concentrated upon smaller regions which appear, under some criterion, to be likely to contain the global optimum. An approach using networks that belongs to this class has been suggested by Robinson (1972). Hartman (1973) examined six variations of the three classes, restricting consideration to the "essentially unconstrained" nonlinear programming problem:

$$\min f(x)$$
$$\text{subject to } x \in S \subseteq E^n$$

where the boundaries of S do not determine the solution. The restriction to E^n is not necessary, so his findings can be applied to the problem at hand.

In general, Hartman found that for the variations used, the first two classes performed better than the third. On difficult problems, though, even the best of the methods will frequently fail to locate the global optimum. Hartman also found that the best results were obtained by methods which do the least random searching.

The subject of avoiding local minima is a fertile field for further research. Approaches for avoiding local minima can be adapted successfully to capitalize on features of a particular problem, and can even be incorporated within the search process itself. One feature of the cyclic search algorithm of Section III.11 is its ability to jump out of the range of a local minimum during a line search. The success of this "escape mechanism" is a function of the initial step size (parameter $\Delta 1$) and the relative deepness of the hollow containing the global optimum as compared to those containing other local minima.

III.14 Application to Multifacility Location-Allocation Problems

III.14.1 Model of Problem

In this section a specific class of location problem, the location-allocation problem, will be considered. Surveys of work on this problem have been presented by Scott (1970) and Cooper (1964), and an extensive bibliography has been accumulated by Lea (1973). A general objective in this class of problem involves determination of the number of new servicing facilities, their location, and optimum allocation of demand points to servicing facilities. In this section, however, it will be assumed that the number of new facilities to be located is known.

A mathematical formulation of this location-allocation problem for the Euclidean norm is:

$$\text{Minimize} \sum_{i=1}^{N} \sum_{j=1}^{M} Z_{ij} W_j [(x_i - a_j)^2 + (y_i - b_j)^2]^{\frac{1}{2}} \quad (3.14.1)$$

$$\text{Subject to} \sum_{i=1}^{N} Z_{ij} = 1 \qquad j = 1, \ldots, M$$

$$Z_{ij} \quad \{0,1\}$$

where: W_j = weights of demand facilities

(a_j, b_j) = P_j are demand point locations

M = number of demand points

N = number of facilities to be located

As exhibited by Cooper (1967) this problem in the plane is neither convex nor concave, and is hence much more difficult to solve than the convex single facility problem. The multifacility location-allocation problem on the sphere is no less difficult.

Two approaches to the problem in the plane are evident, and have been developed over the years. One is via combinatorial programming, recognizing the 0-1 nature of the variables Z_{ij}. A modification of this approach to solve the counterpart multifacility problem on the sphere is the intent of this section. The other major approach, used by Cooper (1967), is via an extension of the single source algorithm. Although such an approach can only obtain local minima, it is very inexpensive with regard to computer time, and is the only efficient approach when solving very large problems.

Mathematical formulation of the spherical multifacility problem is, of course, similar to the planar formulation, with the exception of the different distance measure and the "spherical" equality constraint which restricts solutions to the spherical surface.

$$\text{Minimize} \quad \sum_{i=1}^{N} \sum_{j=1}^{M} Z_{ij} W_j \operatorname{Arccos}(a_j x_i^1 + b_j x_i^2 + c_j x_i^3) \qquad (3.14.2)$$

$$\sum_{i=1}^{N} Z_{ij} = 1 \qquad j = 1, \ldots, M$$

$$(x_i^1)^2 + (x_i^2)^2 + (x_i^3)^2 = 1 \qquad i = 1, \ldots, N$$

$$Z_{ij} \in \{0,1\}$$

where $(a_j, b_j, c_j) = P_j$ are demand points, and all other terminology is as in the planar formulation (3.14.1).

III.14.2 A Planar Algorithm

Kuenne and Soland (1971, 1972) provide a technique for solving the location-allocation problem via branch and bound. A fundamental aspect of the algorithm is sequential solution of a number of single source problems, for which it is possible to find a global optimum in the plane.

One can consider direct application of a great circle metric single facility algorithm in the branch and bound algorithm. However, since only local optimality is obtainable in the general case for the single facility problem on the sphere, one can expect no better than a local minimum for the multifacility problem. Also, as will be seen, the lower bound calculated in the branch and bound algorithm may not be an actual lower bound. Modifications can be made, though, to partially resolve these difficulties.

It is interesting to note that Kuenne and Soland recognized the possible impact of great circle distances and devised an approximation technique for problems covering large regions of the earth. It is based upon a Mercator projection, where distances are computed as rhumb-line map distances and then converted via an approximation to great circle distance (by multiplying the distances by the cosines of the midpoints of the spherical coordinate latitudes). Kuenne and Soland hold that the approximation will work well for most purposes. The large problem solved by Kuenne and Soland covered the contiguous United States, for which a standard Mercator projection is suitable. However, such a projection will fail dramatically if demand points are near the polar regions or in regions larger than a hemisphere.

III.14.3 Modifications for Spherical Problem

Inherent in Kuenne and Soland's algorithm is a presumption that global optimality can be guaranteed for the single facility problem. Single facility sub-problems are solved and the results are used in development of lower bounds, feasible solutions and branching points at each iteration. Noting that the contiguous United States can be located within a disk of diameter $\pi R/2$, where R is the earth's radius, one can be confident of obtaining a global optimum (upon accepting the validity of Conjecture 3.7.1). It is evident that Kuenne and Soland implicitly assumed the validity of the conjecture.

Unfortunately, in the general case global optimality cannot be guaranteed for the single facility problem. This leads to difficulties in determining a lower bound for the branch and bound algorithm. As used by Kuenne and Soland, these bounds consist of the sum of:

1. cost contributions arising from optimally locating r servicing facilities to which more than two demand points have been assigned, and

2. underestimates of costs associated with optimally locating <u>all</u> servicing facilities to service all unassigned demand points in addition to all demand points which had facilities servicing only them.

The underestimates of (2) present no problem since they are based upon inter-demand point distances. No optimization problem is involved. However, since only a local minimum can be guaranteed in general for the location of each of the r servicing facilities in (1), there is no way to discern whether the sum of differential costs between the r global and local solutions exceeds the differential amount between the actual costs of (2) and the calculated underestimate of (2). If it does exceed the differential amount, the calculated lower bound would not be an actual lower bound. In other words, the costs associated with the local solutions may be so "bad" as to overcome the underestimated costs of (2).

Considering the general case, then, it is necessary to modify the Kuenne and Soland algorithm in order to make it adaptable to the sphere problem. It is important to point out, however, that validity of Conjecture 3.7.1 will permit direct application of the branch and bound algorithm (with the spherical single facility location algorithm as a subroutine) when all demand points can be contained within a disk of diameter $\leq \pi R/2$ of their servicing facility.

The procedural steps for the general case required at each iteration follow:

For each tentatively located servicing facility K, check to see if all points allocated to K are contained within a disk of diameter $\leq \pi R/2$ (this determination of the covering circle size is accomplished via the procedures in Chapter IV). If so, solve for the global optimum and use the objective function value for the lower bound contribution. If not, then:

a. solve for the local optimum and use the solution point as an input in determining the branching point and feasible solution.

b. solve the unconstrained problem in E^3 for the Euclidean norm. Use the resulting objective function cost for a lower bound contribution (see Section III.9). As is evident, the lower bounds will not be as tight as in the basic algorithm, so fathoming will not generally occur as early.

Note that once a node is reached where it cannot be concluded that a global optimum is found, no further partitioning from that node will ever result in a global optimum. This is because each branch assigns more demand points to a servicing facility, not less. Hence, the disk, or covering circle, can only increase in size.

The advantage of the preceding approach is that a <u>very good</u> local minimum will eventually result because of the implicit examination of all possible demand point-servicing facility allocations.

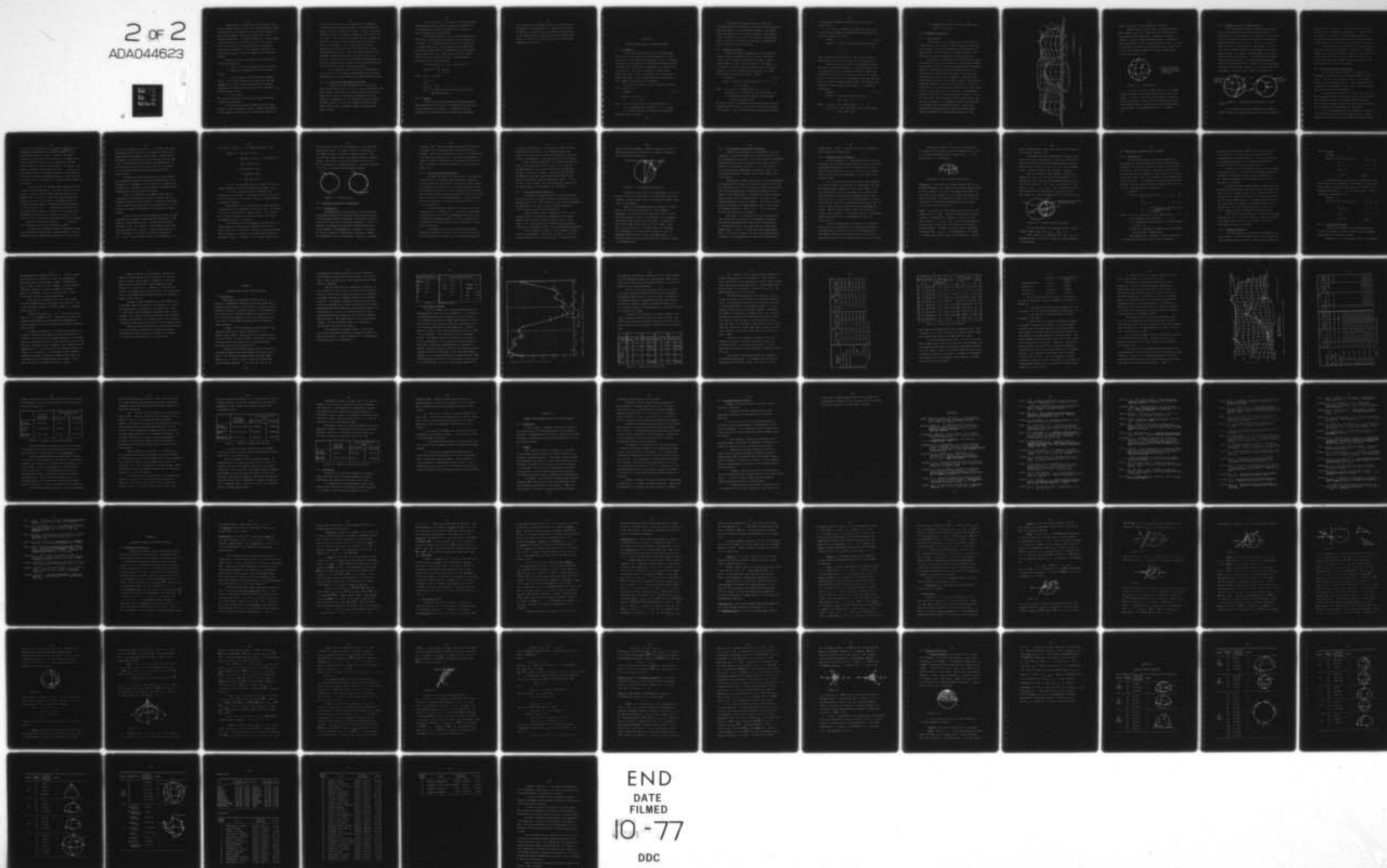
Even though incorporation of the single facility algorithm for the sphere as a basis for the Kuenne and Soland branch and bound approach requires weakening of the lower bounds, it is much better than the Mercator approximation for a problem covering a region as large as the Eurasian-African land mass. As already mentioned, distances between extreme demand points would be distorted significantly in a Mercator Projection using rhumb lines as an approximation to great circle distances. The single source algorithms which have been presented are not affected by the size of the region in determination of inter-demand point distances, nor are they affected by the region's location on the sphere.

III.14.4 Another Approach

As with any branch and bound algorithm, large scale problems are generally not able to be efficiently solved. An increase in servicing facilities (N) along with a corresponding increase in demand points (M) would increase the difficulty level significantly (see Francis and White 1974). So, although a larger number of sources will increase the likelihood of obtaining a final solution which can be asserted to be globally optimal, it is likely the problem will be of such a large scale as to be essentially unsolvable by branch and bound.

AD-A044 623 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/1
LARGE REGION LOCATION PROBLEMS.(U)
1977 D L LITWHILER
UNCLASSIFIED AFIT-CI-77-18 NL

2 OF 2
ADA044623



END
DATE
FILMED
10-77
DDC

Recognizing the difficulty of handling such large scale problems, Cooper has over the years developed a number of heuristics for the Euclidean norm problem in order to get a local minimum. Kuenne and Soland (1971, 1972) coded one and found it extremely efficient. A large number of local minima can be generated rather inexpensively. Since the single facility algorithm for the spherical case can only guarantee a local minimum in general, it would seem practical to use the heuristic for the sphere problem.

The basic steps of the Cooper heuristic used by Kuenne and Soland are:

1. Select N initial servicing facility locations.

2. Assign the M demand points to the closest source.

3. Solve N single facility location problems (results in a global minimum on the plane for each <u>sub-problem</u>; but a local or global minimum on the sphere depending on whether the N subsets are each contained in a disk of diameter $\pi R/2$).

4. Go to step 2.

The iterative process continues until some convergence criterion is achieved.

For practical problems, examination of a large number of local minima may suffice. Cooper (1963) found, as did Kuenne and Soland, that the objective function is fairly flat and hence not very sensitive to changes in location around the global optimum. This property would appear to carry forward to the sphere. In fact, one could reasonably expect the objective function to be flatter on the sphere. On the sphere the single facility problem can even have a constant objective function value for all points in the domain. This occurs when six equally weighted points are placed at the intersection of three orthogonal great circles. Also, the upper bound of πR on distances between points limits the total cost at any location. This is not so for the planar problem since the cost can increase without bound by selecting progressively "worse" locations for the servicing facility.

It is important to note that use of the heuristic can also help in the branch and bound approach by cheaply generating a good initial feasible solution and lower bound.

III.14.5 The Discrete Multifacility Problem

In addition to the formulation of the location-allocation problem given by (3.14.1), other formulations have been studied. One variation involves a discrete solution space and is known as the plant or warehouse location problem. A treatment of this problem is given by Francis and White (1974). It has been studied by ReVelle and Swain (1970), Curry and Skeith (1969), and Shannon and Ignizio (1970).

It is of interest to note that a discrete location/covering problem that closely parallels the objective of the foregoing can easily be handled for the spherical case by using existing algorithms.

An efficient heuristic due to Shannon and Ignizio (1970) will provide "good" solutions for the problem on the sphere. The fundamental data base for the algorithm is a fixed matrix $[a_{ij}]$ of distances/costs derived from interaction between discrete servicing facility locations X_j and demand points P_i. The matrix, of course, would incorporate great circle distances, thereby making it applicable to the sphere problem.

Letting $W_{ji}\rho(X_j, P_i) = a_{ij}$, the objective is:

$$\text{Minimize } Z = \sum_{i=1}^{M} \min_{j \in \theta(X)} a_{ij}$$

where: $\theta(X) = \{j \mid X_j = 1\} \neq \phi$

$$\sum_{j=1}^{M} X_j \leq K$$

$$X_j \varepsilon \{0,1\} \qquad \forall_j$$

W_{ji} is the weight between servicing facility X_j and demand point P_i

III.15 <u>Summary</u>

In this chapter two new heuristic algorithms have been presented for solution of the single facility minisum location problem on the sphere. In addition, several properties of the minisum sphere problem were discussed, and extension of some planar results to the sphere were presented. It was shown that the search for a global optimum may be restricted to the spherically convex hull of the demand points. Application of the solution techniques to solution of a special class of multifacility problems was discussed.

CHAPTER IV

MINIMAX SINGLE FACILITY LOCATION PROBLEMS

IV.1 Introduction

In this chapter minimization of the maximum distance for a single facility location problem on a sphere in the great circle metric is addressed. This is a natural extension of the single facility minisum problem discussed in Chapter III and often is considered to be a more realistic objective. Only the case where all demand point weights are equal will be considered. Two similar geometric approaches to the minimax problem are presented, as are a number of properties of the problem.

As stated in Chapter I, in an ℓ_p norm the problem can be formulated as:

$$\underset{X}{\text{Minimize}}\ Z \qquad (4.1.1)$$

$$\text{Subject to } |X-P_i|_{\ell_p} \leq Z$$
$$i=1,\ldots,M$$

where: X is the location of the servicing facility,

P_i is a demand point location, $i = 1,\ldots,M$, and

Z is (geometrically) the radius of the minimum covering circle or sphere.

As observed by Elzinga and Hearn (1972b) the Euclidean norm (ℓ_2) problem can be solved by convex programming techniques for a unique global optimum. In simplest form, the objective is to find the smallest circle which covers a finite set of points in the plane. Capitalizing on its structure, an efficient geometrical technique can also be used to solve this problem.

IV.2 Problem Formulation

On the unit sphere formulation of the problem is similar, with the exception that the great circle metric is used, and a spherical constraint is added. Given a finite number of points on the surface of the sphere, it is desired to locate a servicing facility on the sphere's surface so that the maximum distance between this point and the given points is minimized. That is:

$$\text{Minimize } Z \tag{4.2.1}$$

$$\text{Subject to Arccos } (a_i x_1 + b_i x_2 + c_i x_3) \leq Z \quad i=1,\ldots,M$$

$$x_1^2 + x_2^2 + x_3^2 = 1$$

where: (a_i, b_i, c_i) is a demand point, $i = 1,\ldots,M$

(x_1, x_2, x_3) is a servicing facility location, and

Z is (geometrically) the radius of the minimum covering spherical cap.

The first constraint represents the distance metric and the second constraint forces the solution to be on the surface of the sphere. The second constraint unfortunately creates a non-convex programming problem, as will be seen in Section IV.3.1.

It is easy to implicitly eliminate the last constraint by using spherical coordinates and formulate the problem in spherical space S^2:

$$\text{Minimize} \quad Z \qquad (4.2.2)$$

$$\text{Subject to Arccos } (a_i \sin\phi\cos\lambda + b_i \sin\phi\sin\lambda + c_i \cos\phi) \le Z \qquad i = 1, \ldots, M$$

$$0 \le \phi \le \pi$$

$$-\pi < \lambda \le \pi$$

where $(\sin\phi\cos\lambda, \sin\phi\sin\lambda, \cos\phi)$ is the location of the servicing facility and (a_i, b_i, c_i) and Z are as before.

Recognizing the one to one correspondence between chord and arc lengths of length $< \pi$ the problem may also be formulated as one in Euclidean E^3 space by adding the constraint which forces the solution to be on the sphere. That is, the objective is to equivalently minimize the maximum chord length (Euclidean norm) rather than minimize the maximum arc length (great circle metric). It would be formulated as:

$$\text{Minimize} \quad Z \qquad (4.2.3)$$

$$\text{Subject to } [(x_1-a_i)^2+(x_2-b_i)^2+(x_3-c_i)^2]^{\frac{1}{2}} \le Z$$

$$x_1^2 + x_2^2 + x_3^2 = 1$$

where: (a_i, b_i, c_i) is a demand point, $i = 1, \ldots, M$

(x_1, x_2, x_3) is the servicing facility in Euclidean three space, and

Z is (geometrically) the radius of a sphere with center at (x_1, x_2, x_3)

IV.3 Fundamental Properties

IV.3.1 Non-convexity

On the plane, the optimum point for the minimax problem exists and is unique. This is intuitively obvious and is shown to be true by Blumenthal and Wahlin (1941), among others. On the surface of a sphere, however, this uniqueness cannot be guaranteed for the general case.

As observed in formulation of the problem, presence of the equality constraint results in a non-convex programming problem. This is most vividly demonstrated by example. Figure 4.1 depicts the objective function plotted as a function of colatitude and longitude for the four point problem using data set D5 of Appendix B. Even with the small number of points, the surface is full of well-defined ridges and valleys, as well as local optima. As will be pointed out, the location of the demand points on the sphere relative to one another is all-important in determining how easy it will be to obtain an optimal solution. It is intuitive that these four points form a global subset and are not containable in a hemisphere (this can be verified by techniques due to Blumenthal (1956)). Solution of problems having demand points which form global subsets cannot be handled by the algorithms presented in this

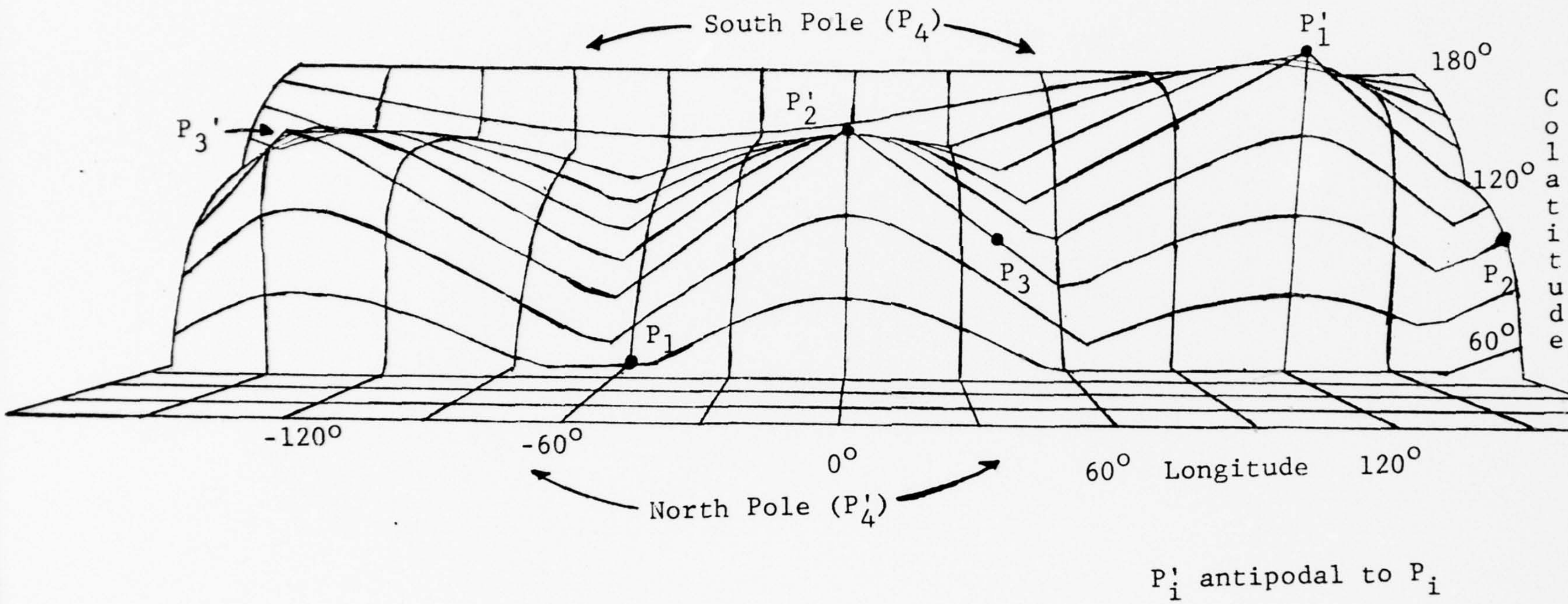
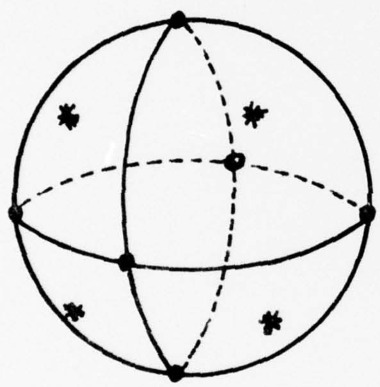


Figure 4.1. Minimax function plot; Data Set D5, Appendix B.

chapter, and hence remain essentially unsolved.

Global subsets can give rise to alternate optima--which cannot occur on the plane due to the convexity of the problem in Euclidean space. Perhaps the simplest example is the case when six demand points are located at the intersection of three orthogonal great circles. In such a case there are eight alternate optima, one in each of the octants of the sphere formed by the intersecting great circles (see Figure 4.2).



* = visible alternate *optima*, with four others on opposite hemisphere.

Figure 4.2. A special case.

A special case also arises when the set of demand points includes two antipodal points and the set is not global. In such a case there are an infinite number of alternative optima. They consist of points on the great circle which forms the equator when the antipodal points are the poles. The minimax distance is, of course, $\pi/2$.

IV.3.2 Comparison with E^n Minimax Problem

The n-dimensional minimum covering sphere problem has been investigated by Elzinga and Hearn (1972b). At first glance, it might appear that a solution of the problem in E^3 can be modified to solve the problem on the sphere using the formulation of (4.2.3). This is so when all demand points are located within a hemisphere, as will be seen, but it cannot be used when the demand points form a global subset. For example, consider again the six points located on the unit sphere S at the intersection of three orthogonal great circles. The minimum covering sphere (MCS) is the sphere S itself. However, the covering sphere determined by an optimal solution to the minimax problem on the sphere's surface is much larger (see Figure 4.3).

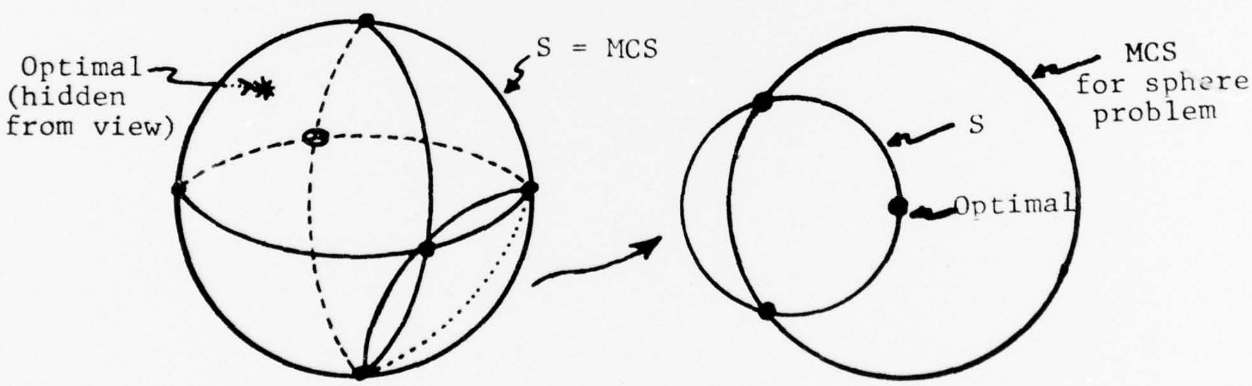


Figure 4.3. Minimum covering sphere for sphere problem.

It is clear that for the problem on the sphere's surface, there is an upper bound for the solution, whereas there is none on the plane. This bound occurs only when all points on the sphere are demand points. Then any point on the sphere is optimal, and the minimax distance is π. In the problem being considered there are a finite number of destinations so the upper bound will never be reached.

Although on the plane two points determine a unique circle for which they are ends of a diameter, on a sphere's surface the points can determine a great circle, a disk, and a spherical circle whose radius is greater than $\pi/2$. As will be seen, when working with global subsets these latter circles are the ones of interest.

IV.4 Difficulties with Global Subsets

If the demand points cannot be contained in a hemisphere, then certainly the minimax distance is greater than $\pi/2$, and thus the covering circle is a non-convex region of radius greater than $\pi/2$. Many difficulties are encountered in attempting to solve such a problem, some of which have already been described.

As mentioned in Chapter II, a solution was offered for the sphere problem by Sylvester (1860), and recognized more recently as such by Blumenthal and Wahlin (1941). The solution technique was actually for the planar problem, but was claimed to be analogous for the sphere.

It is claimed here that the technique outlined by Sylvester will not solve the problem for global subsets. Sylvester was interested in the problem primarily because he found its solution useful in a paper dealing with the approximate representation of $\sqrt{x^2+y^2}$ and $\sqrt{x^2+y^2+z^2}$ by linear functions of x, y and z. Early in the paper, variables are restricted to be positive. It appears that this may have led to the error in concluding that the planar technique is equally applicable to the sphere. For example, Sylvester assumes that one can ". . . reject all those points that are contained within the contour formed by the arcs joining the remaining points, so that the case of points lying at the angles of a convex polygon remain to be studied. . . ."

Sylvester did not consider global subsets and this is where the approach failed. The first step in the technique is to find a circle (disk) which contains all the demand points, and then in following steps progressively decrease the size of the circle. An analogue model is offered via imagination of a rubber band which maintains a circular shape. If ". . . sufficiently stretched over the surface of the sphere to contain all the given points (represented by minute pins' heads given upon it), this band will by its contraction upon the surface of the sphere . . . imitate the method of solution. . . ." Clearly, this procedure will not work for global subsets.

An efficient technique to handle the minimax objective with global subsets awaits further research. Such a problem may be of mathematical interest, but it will only arise in the single facility case. In solving a multifacility problem there is no need to consider any partition of demand points which forms a global subset. Clearly, when the number of servicing facilities is greater than or equal to two, the maximum distance between any demand point and its servicing facility in an optimum solution is bounded by $\pi/2$ merely by placing the servicing facilities at opposite poles.

Although the algorithms presented in this chapter deal only with problems for which the demand points can be contained in a hemisphere, it is instructive to observe the properties of the single facility problem with demand points forming global subsets before continuing.

Once it is known that the demand points are not containable in a hemisphere, then it is obvious that the minimax distance is greater than $\pi/2$. Using the observation that the upper bound on distances on the sphere is π, it is possible to convert the problem to an equivalent "maximin" problem.

Consider the objective of finding the point X_b which maximizes the minimum distance from any demand point. Suppose the optimal distance is b. Consider X_b', the point antipodal to X_b. The distance to any demand point P_i from X_b' is then equal to $\pi-\rho(X_b,P_i)$. Now certainly some P_k is on the boundary of the disk D of radius b centered at X_b. If not, D could be made larger--a contradiction. So for this P_k, $\rho(X_b, P_k) = b$, the maximin distance. Now,

$$
\begin{aligned}
\rho(X_b P_k) = b &= \operatorname*{Max}_{Y} \operatorname*{Min}_{i} (\rho(Y, P_i)) \\
&= \operatorname*{Max}_{X}[\operatorname*{Min}_{i}(\pi - \rho(X, P_i))], \text{ X antipodal to Y} \\
&= \operatorname*{Max}_{X}[\pi - \operatorname*{Min}_{i}(\rho(X, P_i))] \\
&= \operatorname*{Max}_{X}[\pi + \operatorname*{Max}_{i}(\rho(X, P_i))] \\
&= \pi + \operatorname*{Max}_{X}[\operatorname*{Max}_{i} \rho(X, P_i)] \\
&= \pi - \operatorname*{Min}_{X} \operatorname*{Max}_{i}(\rho(X, P_i))
\end{aligned}
$$

So if b, the maximin distance is found, one has the minimax distance, with optimal X_b' being antipodal to X_b.

It is this equivalence which illuminates the difficulty of the problem. The maximin problem, also known in the literature as the "noxious facility" problem remains essentially unsolved.

The problem is of a combinatorial nature requiring consideration of the different disks (less than π in diameter) that can be formed using the demand points. The objective is to find the largest such disk which does not contain any demand points in its interior.

As an example of the foregoing, consider Figure 4.4, which depicts two problems for a degenerate case (on a great circle arc). In Figure 4.4(a) the solution is obvious and can readily be obtained by the algorithm outlined in Sylvester's paper. In Figure 4.4(b), though, there is no disk which will cover all the demand points, and Sylvester's approach will fail. In order to solve it intuitively, it is probably easiest to follow the above procedure. That is, find the largest circle (arc) which contains no demand points. The center of the circle (midpoint of the arc), X^*, is optimum for the maximin problem. The point antipodal to X^*, $X^{*'}$, is optimum for the minimax problem.

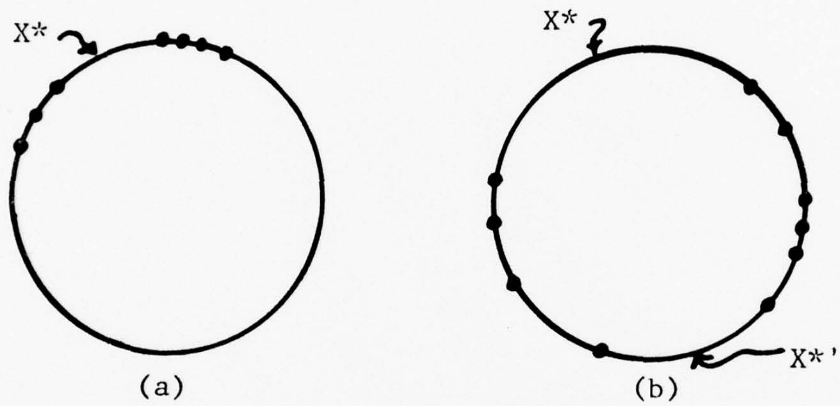


Figure 4.4. Global subsets.

IV.5 Solving the Problem on a Hemisphere

IV.5.1 Introduction

Recognizing that the algorithm of Peirce presented by Sylvester (1860) will solve the problem when all points are containable within a hemisphere, the question naturally arises as to the value of any other approaches. Peirce's approach is designed for solving the problem by hand and is not amenable to programming on a computer. Certainly if the only concern is single facility problems, though, this procedure is quite satisfactory. However, as pointed by Hearn (1971), high speed methods would be desirable for the single facility problem when solving multifacility problems. Due to the combinatorial nature of the multifacility problem (see Section IV.6), any solution procedure is likely to involve decomposition into single facility problems, and usually the resulting number of such problems is quite large.

IV.5.2 The Euclidean Norm Technique

Although new results and approaches to the Euclidean norm problem are continually being developed, currently the most efficient algorithm for solving the planar problem is a geometrical approach due to Elzinga and Hearn (1972a). The algorithm begins with a circle formed by any two points and proceeds to monotonically increase the size of the circle until all points are contained within it. Since there are only a finite number of two and three point circles, it is a finite algorithm.

Elzinga and Hearn (1972b) also provide a solution procedure for the problem in E^n. It involves transforming the Wolfe dual of the convex programming formulation into a quadratic programming problem. A finite decomposition algorithm based on the Simplex method of quadratic programming is developed.

Encouraged by the fact that Elzinga and Hearn's geometric approach is currently the most efficient method for solving the problem in E^2, use of projective geometry was made to capitalize on the technique in order to solve the problem on the sphere. The only additional time required to solve such a problem on the sphere is that due to the transformation to the plane and back to the sphere.

As an alternate approach, it is shown that the solution to the problem in E^3 can be used, via a projection (normalization), to obtain the optimum for the spherical problem. In practice, however, the first approach would be preferred since the planar algorithm has been coded and is more efficient (Elzinga and Hearn 1972b). It must be kept in mind, though, that both approaches require that all demand points be containable in a hemisphere.

IV.5.3 <u>The Stereographic Projection</u>

The geometric algorithm for E^2 can be effectively utilized to solve the sphere problem through use of a stereographic projection. This projection, most likely discovered by Hipparchus (circa 160-125 B.C.), was discarded for centuries when the world was decreed flat.

Kreysig (1968) credits Lagrange with first establishing that the stereographic projection is the unique mapping which preserves circles from the sphere to the plane. It is this property upon which one can capitalize, recognizing the geometric basis of the Elzinga-Hearn algorithm.

Cotter (1966) (cf. Hilbert 1952) provides a simple proof of orthomorphism. Referring to Figure 4.5 he established the similarity of ΔXab and ΔXAB. He then observed that for similar triangles, cones of rotation are similar and the projection of a circle on the sphere of diameter AB is a circle of diameter ab.

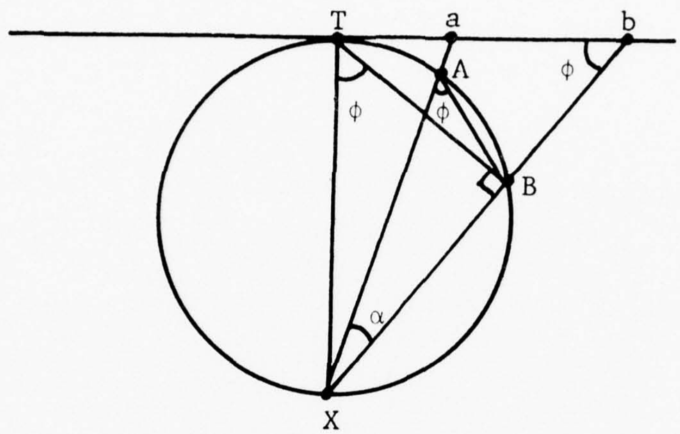


Figure 4.5. Stereographic projection.

Through the projection, the correspondence is one to one onto. That is, to each circle C on the plane, there corresponds a unique small circle C' of spherical diameter less than π on the sphere.

Now, if all demand points are contained in a hemisphere, and the projection point is in the opposite hemisphere, it follows via the projection that if a point X is contained in a circle C on the plane, its corresponding inverse image point X' on the sphere is contained in the inverse image small circle C' on the sphere. Note that this is NOT true if all points cannot be contained in a hemisphere or if the projection point is not in a hemisphere which doesn't contain the demand points. A suitable projection point would simply be a point antipodal to any demand point.

IV.5.4 Stereographic Projection/E^2 Technique

Using the above properties and observation to best advantage, a parallel problem can be constructed in E^2 via the stereographic projection. This single facility unweighted minimax problem can then be solved in the plane using the efficient Elzinga-Hearn algorithm. In E^2, the optimum circle C_p^* is unique (Rademacher and Toeplitz 1957). Via the projection, C_p^* has a unique inverse image small circle on the sphere.

Requiring all demand points to be within a hemisphere, and appropriately choosing the projection point in the opposite hemisphere, note that the spherical small circle C_s^* contains all the demand points whenever C_p^* contains all the images of the demand points. Certainly no smaller spherical circle contains all the demand points. C_s^*, then, is the minimum covering circle on the sphere. Its center is optimum for the spherical minimax problem since at this point the maximum distance to some P_i, $i=1,\ldots,M$, is obtained, and this distance can be no smaller.

Once again, it is important to note the necessity that all points be contained in a hemisphere, otherwise the inverse transformation's small circle corresponding to the optimum circle in the plane will not contain all demand points. In fact, it would not contain _any_ demand points. For similar reasons, the pole for projection must be exterior to the spherically convex hull containing all demand points. That is, the pole must be in a hemisphere which contains no demand points.

IV.5.5 Normalization/E^3 Technique

In the following it is established that a solution technique for the equally weighted single facility Euclidean norm minimax problem in E^3 can be advantageously used to solve the same problem on the unit sphere S if all demand points P_i are on an open hemisphere of S and great circle distances apply.

It is obvious that if any two points are antipodal, or if the P_i cannot be contained in a hemisphere, the minimum covering sphere (MCS) will be the unit sphere S, and the optimum point would be the center of S. Such a situation precludes solution by the technique which follows. As will be seen, a crucial part of the procedure is to project from the center of S through the center of the MCS, X* to get a unique point on the surface of S (normalization of X*). This certainly cannot be done if the center of S and the MCS coincide. A rudimentary method of determining whether all points are in a hemisphere was given in Section III.4.3.

Prior to establishing the theorem it is useful to observe that on a hemisphere, the correspondence between the small great circle arc and the chord determined by two distinct points via the projection of points of the chord onto the arc from the sphere center is one to one onto.

Referring to Figure 4.6, on an open hemisphere it is easily seen that as X increases monotonically so does A; the same relationship holds between x and a. It also follows that $a = R\ \mathrm{Arcsin}(x/R)$.

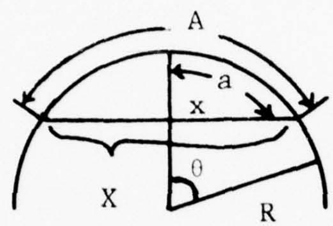


Figure 4.6. Arc and chord relationship.

<u>Theorem 4.1</u> Given M equally weighted demand points P_i on an open hemisphere of S, suppose the Minimum Covering Sphere in E^3 is obtained. Let X_s^* be the unique projection from the center of the sphere S to the surface through X_0^*, the center of the MCS. Considering the points P_i, X_s^* is the optimum for the minimax distance problem for the great circle metric on the surface of S.

<u>Proof</u>: Certainly one of the demand points will be on the boundary of the E^3 MCS. Otherwise the MCS could be made smaller, a contradiction. Suppose P_j is such a point (see Figure 4.7). Let $|X_o^* - P_j|_{\ell_2}$ be the optimum minimax distance in E^3, where X_o^* is the center of the minimum covering sphere. Consider the chord from P_j through X_o^* of length $2|X_o^* - P_j|_{\ell_2} = X$. Corresponding to this chord is a unique small great circle arc of measure A. Letting

X_s^* be the projection of X_o^* to the surface of the sphere it is seen that $\rho(X_s^*,P_j) = \frac{1}{2}A$.

The proof that X_s^* is optimum for the great circle metric on S is by contradiction. Let X_s' be a point on S distinct from X_s^* such that its maximum distance from any other point is $\rho(X_s',P_k)$, where P_k is one of the M demand points. Further, suppose that $\rho(X_s',P_k) < \rho(X_s^*,P_j)$.

Now, as observed, corresponding to $\rho(X_s',P_k)$ is length $|X_o' - P_k|_{\ell_2}$ via the inverse correspondence. It follows that $|X_o' - P_k|_{\ell_2}$ is the maximum length from X_o' to any of the points P_i. For, if not, there exists P_1 such that $|X_o' - P_1|_{\ell_2} > |X_o' - P_k|_{\ell_2}$. Due to properties of the Arcsin function this implies that $\rho(X_s',P_1) > \rho(X_s',P_k)$. This cannot be, since by assumption P_k is further from X_s'.

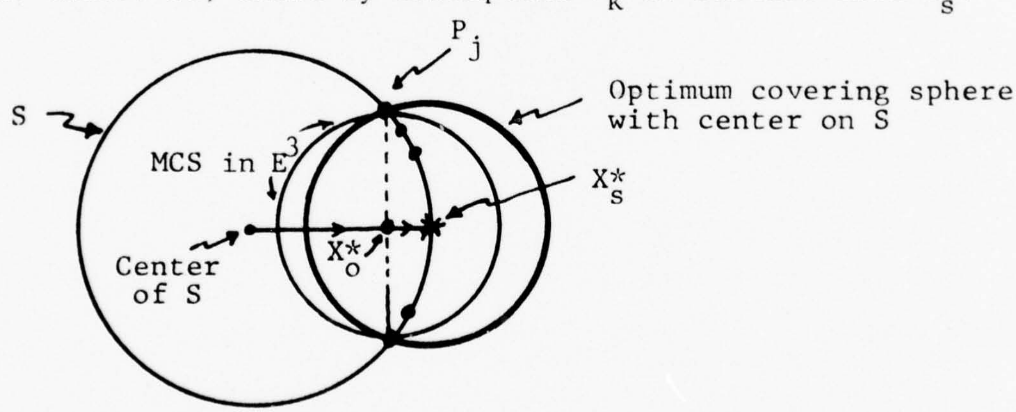


Figure 4.7. Normalization/E^3 technique.

In the same manner, the assumption that $\rho(X_s',P_k) < \rho(X_s^*,P_j)$ implies that $|X_o' - P_k|_{\ell_2} < |X_o^* - P_j|_{\ell_2}$.

This results in a conclusion that $|X_o' - P_k|_{\ell_2}$ is a maximum distance strictly less than the minimax distance, a contradiction.

IV.6 Application to Multifacility Problem

IV.6.1 Introduction

As mentioned earlier, the primary incentive for developing an efficient algorithm for the single facility problem is its possible usefulness in solving a multifacility problem. In this section a particular type of multifacility problem is considered in which there is no interaction between facilities to be located, and all demand point weights are equal. The problem in E^2 was briefly discussed by Hearn (1971) and also studied by Loginov (1969) for the E^n case. The mathematical formulation is:

$$\text{Minimize} \quad V$$

$$\text{subject to} \quad \max_{P_i, X_j} \{Z_{ji} \cdot |X_j - P_i|_{\ell_2}\} \leq V \qquad i = 1, \ldots, M; \; j = 1, \ldots, N$$

$$\sum_{j=1}^{N} Z_{ji} \geq 1 \qquad j = 1, \ldots, N$$

$$Z_{ji} = \begin{cases} 1 & \text{if facility } X_j \text{ interacts with demand point } P_i \\ 0 & \text{otherwise} \end{cases}$$

where: X_j is the location of a servicing facility, $j = 1, \ldots, N$

P_i is a demand point location, $i = 1, \ldots, M$

V is (geometrically) the radius of a disk centered at an X_j, and

M and N are the number of demand points and servicing facilities, respectively.

The problem can be stated as: Given a finite set A of points in Euclidean n-space, and an integer N, determine N n-dimensional spheres such that each point of A is contained in at least one of the spheres and the maximum sphere has minimum diameter. It can be considered as a minimax counterpart to the multifacility minisum location-allocation problem of Section III.14.

Loginov's solution procedure will solve the problem for small N and M in E^3. Hearn (1971) presents an implicit enumeration technique which will solve the problem in E^2. Neither technique is very efficient, but at least a procedure is available.

The multifacility minimax problem presents the same difficulties encountered in the multifacility minisum problem. Due to its combinatorial nature, current solution procedures cannot handle more than 3 or 4 servicing facilities and 20 or more demand points. The number of partitions for any given M and N is S(M,N), the Stirling number of the second kind. In the particular case of N=2, $S(M,2) = 2^{M-1} - 1$, which quickly increases for moderate size M.

The intent of this section is to show that Loginov's procedure for E^3, or any more efficient technique that may be developed in the future, can be used to solve the counterpart problem on the sphere.

IV.6.2 Problem Formulation

For the problem on the sphere, the formulation is quite similar, except that great circle arcs are used, and the servicing facilities are restricted to the surface of the unit sphere:

$$\text{Minimize} \quad v$$

$$\text{Subject to } \operatorname{Max}\{\rho(X_j, P_i) \cdot Z_{ji}\} \leq v \qquad i = 1,\ldots,M;\ j = 1,\ldots,N$$

$$\sum_{j=1}^{N} Z_{ji} \geq 1 \qquad i = 1,\ldots,M$$

$$|X_j|_{\ell_2} = 1 \qquad j = 1,\ldots,N$$

$$Z_{ji} \varepsilon \{0,1\} \qquad \forall\, i,j$$

Recognizing that as long as all demand points are contained within a hemisphere, it follows that the maximum great circle arc length corresponds to the length of the maximum chord, which is ≤ 2 for the unit sphere. Thus, an equivalent formulation is to:

$$\text{Minimize} \quad v$$

$$\text{Subject to } \operatorname{Max}\{|X_j - P_i|_{\ell_2} \cdot Z_{ji}\} \leq v \qquad i = 1,\ldots,M;\ j = 1,\ldots,N$$

$$\sum_{j=1}^{N} Z_{ji} \geq 1 \qquad i = 1,\ldots,M$$

$$|X_j|_{\ell_2} = 1 \qquad j = 1,\ldots,N$$

$$Z_{ji} \varepsilon \{0,1\} \qquad \forall\, i,j$$

IV.6.3 A Solution Technique

It will now be shown that any procedure which solves the problem in E^3 can be used to find an optimum solution for the spherical problem.

Suppose $N \geq 2$. As mentioned earlier, the optimum minimax distance is bounded above by $\pi/2$. In fact, a minimax distance of $\pi/2$ would occur only in a pathological case where there are an infinite number of demand facilities (a great circle arc, for example). In this instance there are, likewise, an infinite number of alternate optima. For example, any two antipodal points located on the great circle will be optimum points.

For practical purposes, then, it can be assumed that the minimax distance is strictly less than $\pi/2$. This is the case when there are a discrete number, M, of demand points and $N \geq 2$.

Suppose all points P_i, $i = 1,\ldots,M$ are located on a sphere S. S is embedded in E^3. Capitalizing on this fact, consider the problem as unconstrained and employing an existing algorithm for the counterpart multifacility minimax problem in E^3.

Solving the problem in E^3, the solution consists of N spheres, the centers of which are servicing facilities optimally located so that the maximum distance to any demand point P_i is minimized and all P_i are serviced by at least one facility. The union of the N spheres, then, contains all the demand points. It is obvious that the solution need not be unique, i.e., sometimes it is possible to find alternate systems of N spheres covering all demand points such that the diameter of the maximum sphere is the same. In any case, of course, the minimax distance is unique.

Given an optimal set of N spheres, project the centers of these spheres to the surface of S. Since no point is greater than distance 2 (diameter of the unit sphere) from any other point, the N spheres are each certainly no larger than S. Likewise, as mentioned previously, the maximum great circle distance from a projected point to any demand point contained within the respective sphere is less than $\pi/2$.

Through a similar argument as in the single facility minimax problem for points on a hemisphere it is seen that monotonicity is preserved in the projection. That is, the largest sphere generates the largest great circle distance. So, using the projected points as the N minimax locations on the sphere, the largest great circle distance is found by projecting from the center of the largest sphere.

This great circle distance is the smallest possible. For, if there were a smaller one, through the inverse transformation there would exist a sphere with diameter less than the minimax diameter, a contradiction.

CHAPTER V

COMPUTATIONAL RESULTS AND CONCLUSIONS

V.1 Introduction

This chapter accomplishes three objectives. It provides a fundamental comparison of the two algorithms in Chapter III with regard to their ability to reach a global optimum (robustness), provides insight to the properties of the single facility minisum problem on the spherical surface, and provides evidence of the error which arises due to the use of a Euclidean assumption when solving large region problems.

A number of test problems have been solved by each algorithm, varying the level of difficulty. Two example "large region" problems from the literature have also been solved. In each problem, verification of global and local optima was made by use of three dimensional plotting and an exhaustive grid search.

Comparisons are made between solutions found via a Euclidean assumption and those found using a great circle metric. The Hyperboloid Approximation Procedure (HAP) (Eyster, et al. 1973) is used to solve the problems for the Euclidean assumption. It is seen that even in the cases where the optimum location is not very different, the difference in objective function values can be significant. Other conclusions are also drawn based upon computational experience.

The two single facility algorithms of Chapter III were coded in Fortran IV, and all problems were solved on the IBM 370/158J system using double precision arithmetic. No attempt was made to determine sensitivity of parameter selection. For example, judicious selection of the step sizes could significantly affect the ability of the Cyclic Meridian Search (CMS) algorithm to reach global optima, but confirmation is a subject for future research. Also, neither algorithm was coded with the intent of achieving high efficiency. The primary goal was to reach a global optimum. For more efficient codes it is necessary to incorporate acceleration procedures.

The parameters provided in Table 5.1 were used for solving the example problems in Appendix B under the great circle metric. Refer to Chapter III for descriptions and definitions of each parameter.

Planar Projection	Cyclic Meridian Search	
BOUND = $\pi/6$	$\Delta 1 = 10^{-1}$	$\alpha = 2$
EP1 = 10^{-6}	ΔINC = 10^{-1}	ITBOUN = 40
EP2 = 10^{-6}	ΔMIN = 10^{-6}	LRANBO = 4
N1 = 50	$\varepsilon = 10^{-9}$	EP1 = 10^{-6}
	S = 1	EP2 = 10^{-6}
		N = 5

Table 5.1. Parameter values.

V.2 Some Example Problems

The first example is a minisum problem using the demand points in Data Set D6 of Appendix B. It consists of 13 equally weighted points located on a great circle arc. It is informative since its simple nature makes it easy to see the effect that different starting solutions have on the iterative procedures of both algorithms. Figure 5.1 depicts a graph of the objective function values as a function of longitude. The latitude is 0^{o} for all demand points. The demand points are indicated in the figure, as are the locations of all starting solutions except S5. Note that a convergent algorithm would be expected to achieve the global optimum (0,20) for any starting solution between -60^{o} and 56^{o} longitude. Both algorithms behave properly, with a surprising result arising from use of CMS. Using starting solution S2, one would expect to get stuck at a local minimum, yet due to the internal features of

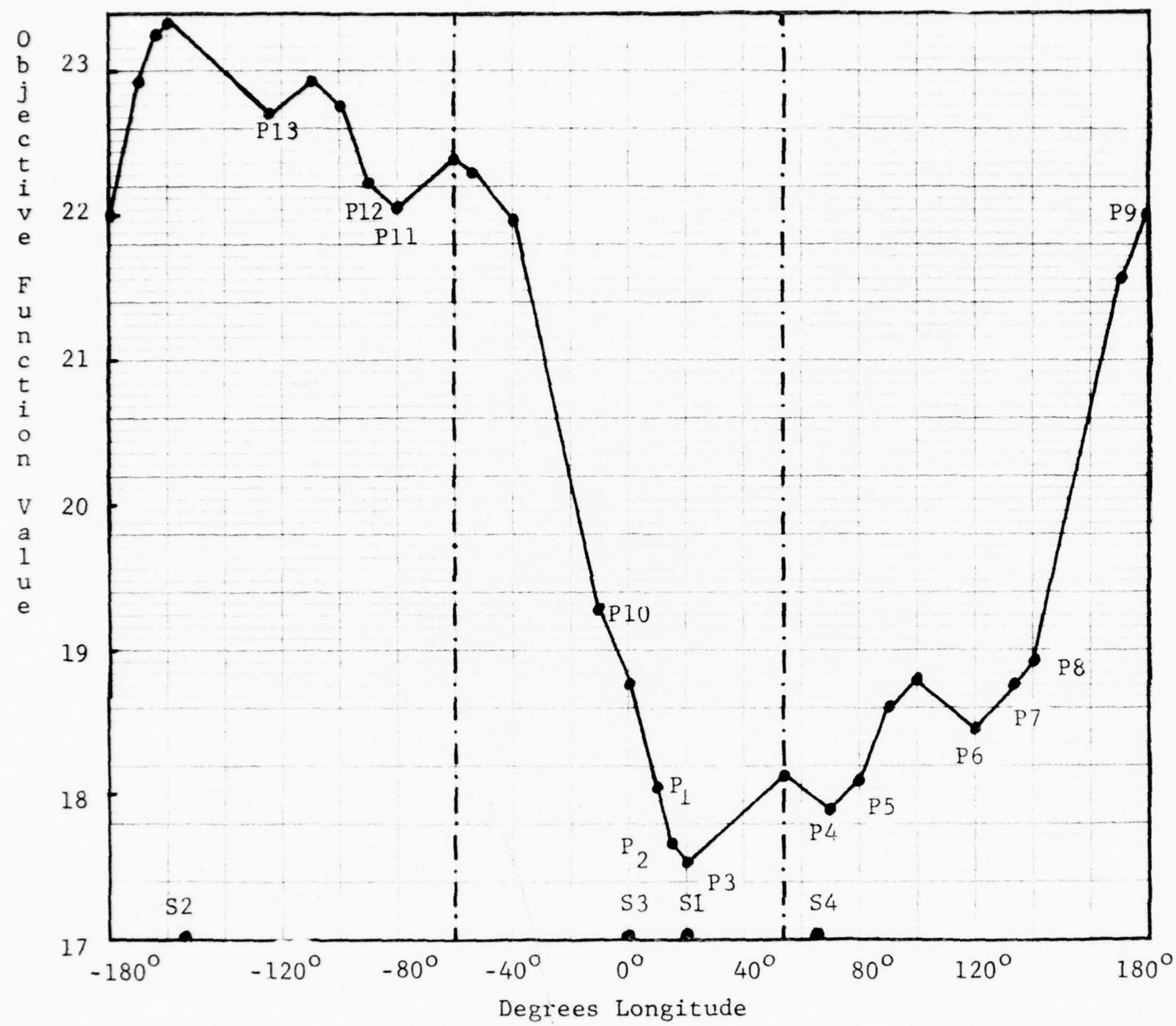


Figure 5.1. Minisum function plot; Data Set D6 of Appendix B.

the algorithm (discussed in Chapter III) the search escapes the local minimum's region of attractiveness, jumps across another one, and stops at the global optimum. This feature seems to occur rather frequently.

Also, note that when starting solution S5 is used, the Planar Projection Algorithm (PPA) reaches the global optimum while CMS stops at a local optimum. Curiously, in all example problems attempted this was the only instance in which PPA achieved a global optimum when CMS did not. It is enough, however, to encourage one to use each algorithm as a check on the other.

In Table 5.2 the results of trying a number of different starting solutions on Data Set D6 are listed. Note that in this problem a projected centroid starting solution (S4) results in a local minimum, regardless of the technique used.

	PPA				CMS			
Starting Point	# Iter	Stops at	Obj Fn Value	Type Soln	# Iter	Stops at	Obj Fn Value	Type Soln
S1(0,20)	1	(0,20)	17.52	Global	1	(0,20)	17.52	Global
S2 (0,-155)	3	(0,-124)	22.69	Local	2	(0,20)	17.52	Global
S3 (0,0)	2	(0,20)	17.52	Global	1	(0,20)	17.52	Global
S4 (0, 62.27) (Proj. Cent.)	2	(0,70)	17.91	Local	1	(0,70)	17.91	Local
S5 (90,90)	9	(0,70)	17.52	Global	6	(0,70)	17.91	Local

Table 5.2. Results for Data Set D6.

Now consider a three equally weighted demand point problem using Data Set D1 of Appendix B. This problem tests the ability of the algorithms to find a global optimum over a definitely non-convex surface. A three dimensional function plot of the problem is found in Figure 3.1. The global optimum is at P3, (30,-60).

The algorithms stopped at the same point, as can be seen in Table 5.3, for all but one of the starting solutions. An interesting observation is that when one starts at the antipodal point to the global optimum, P3', both searches still end up at the global optimum. Concerning the number of iterations, in general CMS is more efficient. Although not concerned with efficiency in terms of execution times in this research, it is interesting to note that CMS takes, on the average, about half the time as PPA. There are situations, though, when CMS is slower than the other approach, particularly when the search begins on or near a ridge.

The next set of examples consists of ten small problems for which a projected centroid was used as a starting solution. As can be seen in Table 5.4, both algorithms successfully found the global optimum in all but one case.

This group of problems gave rise to a number of interesting observations. For example, Data Set D12 was handled more efficiently by PPA. Data Set D9, on the

Starting Point	PPA				CMS			
	# Iter	Stops at	Obj Fn Value	Type Soln	# Iter	Stops at	Obj Fn Value	Type Soln
Proj. Centroid	4	(30,-60)	3.702	Global	6	(30,-60)	3.702	Global
North Pole (90,0)	5	(30,-60)	3.702	Global	5	(30,-60)	3.702	Global
South Pole (-90,0)	9	(0,37.45)	4.171	Local	7	(30,-60)	3.702	Global
P2 (0,36.86)	1	(0,37.45)	4.171	Local	1	(0,37.45)	4.171	Local
P1 (0,180)	1	(0,180)	4.507	Local	1	(0,180)	4.507	Local
P1' (0,0)	6	(30,-60)	3.702	Global	6	(30,-60)	3.702	Global
P3' (-60,120)	9	(30,-60)	3.702	Global	3	(30,-60)	3.702	Global

Table 5.3. Results for Data Set D1.

Data Set	# Pts	Type Soln	No. Iterations		Proj. Centroid	Obj Fn Value	Optimum Soln	Obj Fn Value
			PPA	CMS				
D3	7	Global	3	4	(25.2,27.6)	6.46	(25.8,21)	6.15
D7	3	Global	3	2	(32.3,0)	2.8	(54.9,0)	2.59
D8	3	Global	3	2	(32.3,-60)	2.8	(54.9,-60)	2.59
D9	3	Global	6	1	(63.6,75)	3.64	(80.6,75)	2.08
D10	3	Local	7	3	(0,90)	4.11	(-5.3,90)	4.11
D11	3	Global	3	5	(27.4,71.5)	2.78	(50,80)	2.6
D12	4	Global	2	7	(70.8,37.8)	1.44	(80,45)	1.3
D13	4	Global	6	3	(49.3,110)	4.41	(44.8,50.3)	4.23
D14	4	Global	5	3	(45.9,88.3)	3.79	(52.1,100)	3.76
D15	6	Global	1	1	(0,0)	9.42	(.97,0)	9.42

Table 5.4. Results for ten problems.

other hand was dispensed with in one iteration by CMS. This is because in the first iteration the search occurred along the meridian upon which the optimum solution exists.

In the large majority of cases, the first iteration makes a major move toward the global (or local) optimum, and the remaining iterations involve small steps which continually, but slowly, improve the objective function value until a stopping criterion is satisfied. In order to illustrate this property of the algorithmic search pattern, consider Data Set D7 as solved by PPA:

	Location	Objective Function Value
Starting Solution:	(32.3,0)	2.797699
Iteration 1:	(53.97,0)	2.59099
Iteration 2:	(54.84,0)	2.590403
Iteration 3:	(54.86,0)	2.590402

Certain observations about some of the ten Data Sets are of interest in light of comments and results in Chapter III.

1. Data Sets D3 and D12 are containable in a disk of diameter $\pi/2$, and can be considered via Conjecture 3.7.1.

2. Data Set D14 illustrates Fagnano's Result extended to the sphere.

3. Data Sets D7 through D11 demonstrate aspects of Steiner's Problem on the sphere. It can be easily verified that the optimum solution is at a point which forms angles of at least 120° with the three vertices of the spherical triangles formed by the demand points. The global optimum for Data Set D11 occurs at a vertex, which has an angle greater than 120°. There are two alternate global optima for Data Set D10, but neither algorithm reached one by using the projected centroid as a starting solution. The alternate optima occur at the vertices (-30,20) and (-30,160). The local optimum, at which iterations stopped, is the interior point of the spherical triangle which forms angles of precisely 120° with lines drawn to the three vertices.

4. Data Set D15 has a constant objective function value. Every point on the sphere is optimal.

Now consider a problem using Data Set D16 of Appendix B. This six unequally weighted demand point problem was solved by both algorithms using a variety of starting solutions consisting of the optimum found via a Euclidean assumption, the projected centroid, all demand points, the antipodal point to the projected centroid and antipodal points to all demand points. By referring to Figure 5.2, one can see intuitively the expected "goodness" or "badness" of the respective starting solutions relative to the location of the global optimum.

The computational results are exhibited in Table 5.5. For the various starting solutions used, CMS performs better overall than PPA in reaching the global optimum. Also, in execution time CMS achieved the solution twice as fast, on the average, as PPA.

Of particular interest are the results when starting solutions S6 and S8 are used. Although they are local minima, CMS successfully escapes their grasp and reaches the global optimum.

Note that in this problem the projected centroid is a good starting solution, and is not too far from the optimal solution. In fact, it is significantly better than what would have been achieved if the problem had for some reason been solved with a Euclidean assumption. Table 5.6

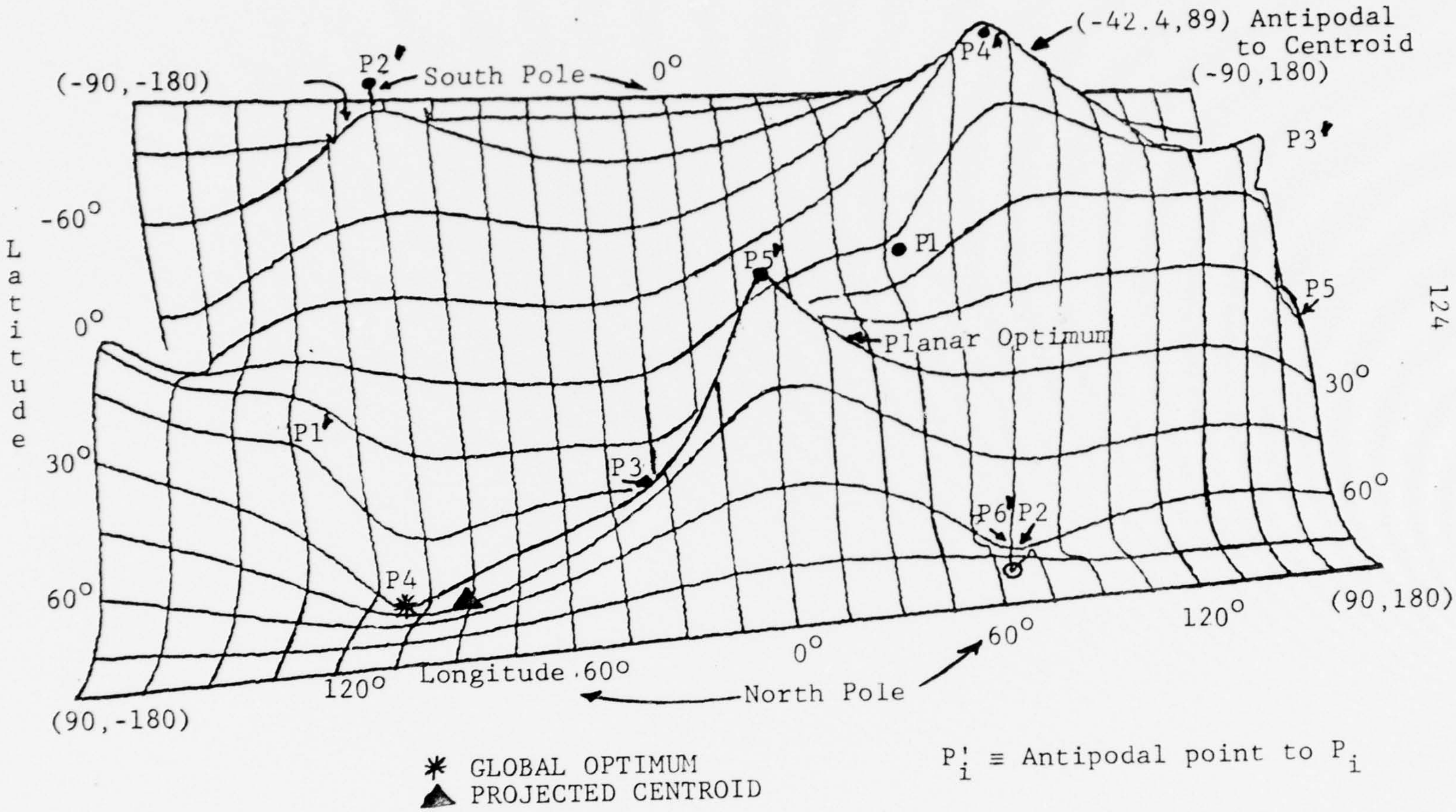


Figure 5.2. Function plot of Data Set D16; Appendix B.

Start Point	PPA				CMS			
	# Iter	Stops at	Obj Fn Value	Type Soln	# Iter	Stops at	Obj Fn Value	Type Soln
S1 (47.6,-91) Proj. Cent.	4	P4(25,-115)	20.5698	Global	6	P4(25,-115)	20.5698	Global
S2 Antipode to S1	6	P5(-30,175)	21.3018	Local	6	P4(25,-115)	"	Global
S3 Planar Optimum	10	P3(15,-20)	21.0766	Local	7	"	"	"
S4 P1(-12,48)	30	P3(15,-20)	21.0766	Local	9	"	"	"
S5 P1'	3	P4(25,-115)	20.5698	Global	9	"	"	"
S6 P2(65,75)	1	P2(65.75)	21.0789	Local	9	"	"	"
S7 P2'	4	P5(-30,175)	21.3018	Local	9	"	"	"
S8 P3(15,-20)	2	P3(15,-20)	21.0766	Local	8	"	"	"
S9 P3'	2	P4(-30,175)	21.3018	Local	8	"	"	"
S10 P4(25,-115)	1	P4(25,-115)	20.5698	Global	2	"	"	"
S11 P4'	31	P3(15,-20)	21.0766	Local	2	"	"	"
S12 P5(-30,175)	1	P5(-30,175)	21.3018	Local	2	P5(-30,175)	21.3018	Local
S13 P5'	5	P2(65,75)	21.0789	Local	9	P4(25,-115)	20.5698	Global
S14 P6(-70,-110)	8	P5(-30,175)	21.3018	Local	7	P5(-30,175)	21.3018	Local
S15 P6'	3	P2(65,75)	21.0789	Local	7	P4(25,-115)	20.5698	Global

Table 5.5. Results for Data Set D16.

presents the results for the projected centroid, the optimum for the sphere, and the optimum found by using a Cartesian coordinate system in E^2.

	Location (Latitude, Longitude)	Objective Function Value	
		Great Circle	Euclidean
Optimum for Sphere	(25,-115)	20.57	37.486
Optimum for E^2	(9.44,21.02)	22.195	25.214
Projected Centroid	(42.4,-91)	20.90	35.434

Table 5.6. *Sphere vs. plane, Data Set D16.*

Using the Euclidean assumption for this particular problem will result in an objective function value error of about 18% and a location error of over <u>10,000 miles</u>!

The next example required the algorithms to solve a problem in the polar regions. Data Set D17 of Appendix B includes the coordinates, in latitude and longitude to the nearest degree, of nine experimental stations and other sites in Antarctica. Using the minisum criterion and assuming equal weights, an optimal solution was found by both algorithms. The projected centroid was used as a starting solution. PPA used three iterations while CMS used four iterations to reach the optimum at (-87.79,89.78).

With the foregoing observations in hand, two problems from the literature were examined. Both involved solution of a single facility minisum problem by using a Euclidean assumption and letting latitude and longitude represent the Cartesian coordinates.

First, consider an example problem used by Kuhn and Kuenne (1962). The data (Data Set D18 of Appendix B) consists of 24 cities of the Ukrainian Soviet Socialist Republic. The weights are equal to the proportion each had of the population of the 100 largest Russian cities (as of 1961). The coordinates are degrees of north latitude and east longitude correct to the nearest full degree. The latitude of Kharkov is in error by about nine degrees, but since Kuhn and Kuenne evidently used the published coordinates to solve the problem, they were also used for this example.

Before examining the results, it is important to point out that the points could easily be fit into an 800 by 1000 mile rectangle, and certainly into a disk of diameter less than $\pi R/2$, where R is the radius of the earth. By Conjecture 3.7.1, one can guarantee a global optimum. Both algorithms achieved the global optimum using the projected centroid as a starting solution.

In this particular problem there is no significant effect on location of the optimal solution relative to using either the great circle metric or Euclidean norm. However, there is a definite effect on the value of the objective function (see Table 5.7). This would be of significant interest in the area of logistics, and especially budgeting, in any attempts to estimate transportation requirements/costs.

	Location (Latitude, Longitude)	Objective Function Value	
		Great Circle	Euclidean
Optimum for Sphere	(47.795,35.2402)	.01293768	.01669464
Optimum E^2 (HAP)	(47.707,35.104)	.01295154	.01663096
Optimum E^2 (Kuhn/Kuenne)	(47.6,35.32)	.01296313	.0166528

Table 5.7. Sphere vs. plane, Data Set D18.

It is easy to verify that the error in optimal location is about 9.26 miles, while the objective function value error is off by about 22%.

The last example is due to Chapelle (1969). He considered the 49 continental states plus the District of Columbia and their corresponding output in first class letter mail in the year 1965. The output volume in pounds was assumed to originate from the capital of each state. Latitude, longitude and weight for each point location are found in Data Set D19 of Appendix B. Chapelle's objective was to optimally locate a central facility for the postal system, assuming plane geometry.

Although the region is larger than in the previous example, it still can be contained in a disk of diameter less than $\pi R/2$. Both algorithms of Chapter III reached the global solution, using the projected centroid as a starting solution, taking seven iterations of CMS and three iterations of PPA to meet stopping criteria. It took 65% less time to solve the problem by CMS.

Referring to Table 5.8, one can compare the results obtained by either a Euclidean norm or great circle metric approach. The error in location is only 30 miles, yet the error in objective function value is over 20%.

	Location (Latitude, Longitude)	Objective Function Value	
		Great Circle	Euclidean
Optimum for Sphere	(40,83)	780,065	987,413
Optimum for Plane	(39.65,-83.45)	781,635	986,878

Table 5.8. Sphere vs. plane, Data Set D19.

V.3. Conclusions

Based upon the experience obtained in solving a number of single facility minisum problems using both algorithms, the following conclusions can be drawn:

1. Comparing the two algorithms in their rudimentary form, the Cyclic Meridian Search (CMS) algorithm is, in general, faster than the Planar Projection

Algorithm (PPA). There is always the possibility of getting stuck on a ridge and zigzagging, but the integral use of random search directions helps avoid these difficulties.

2. PPA is not as likely to reach the global optimum as CMS. Other than the observation that CMS can "jump out" of local minima regions and accelerate in a search direction, there is no apparent a priori reason for its success.

3. No case of divergence was observed for either algorithm in any problem attempted. In all cases at least a local minimum was obtained.

4. The projected centroid is a good starting solution, but will not necessarily result in finding the global optimum.

5. A Euclidean assumption for even moderately large regions (800 by 1000 miles) can lead to significant error in objective function values. This is because a planar assumption overestimates actual distance. Also, significant error in location of the optimal facility can occur in problems covering regions as large as a hemisphere.

CHAPTER VI

SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

VI.1 Introduction

In this chapter a summary of research performed on the single facility location problem for large regions is presented. Recommendations for future research are included to assist in identifying topics which would extend the current research effort.

VI.2 Summary

This research presents an extension of the generalized Weber problem and of the related problem of minimization of maximum distance. The primary objectives were to investigate the effect of a great circle metric on the problems of interest, and to develop solution procedures to solve them. The results will permit greater realism and accuracy in solving large region location problems for which a Euclidean (planar) assumption is inappropriate.

In Chapter I the necessity of considering location problems under a great circle metric is discussed, along with possible applications of the research. Chapter II then presents a literature survey of previous research concerning large region location problems.

In Chapter III the single facility generalized Weber problem is studied, where the goal is to locate a new facility on the sphere relative to several weighted demand facilities such that the total cost of transportation is minimized. Cost is considered to be strictly a function of great circle distance and demand point weights.

Different formulations and properties of the problem were discussed, along with two new iterative algorithms for solving it. Both algorithms were programmed and optimal solutions were obtained for several problems. Application of the techniques to solving a multifacility location-allocation problem is also discussed.

In Chapter IV the objective of locating a facility to minimize the maximum distance to any demand point under a great circle metric was considered. It is assumed that the number of demand facilities is finite and their weights are all equal. Two approaches to the problem are presented, both of which capitalize on existing algorithms to solve the Euclidean norm problem in E^2 or E^3. Properties of the problem and application of the single facility techniques to solution of a certain multifacility minimax problem are presented.

Chapter V presents results of applying the algorithms in Chapter III to a number of example problems. Comparative effectiveness in reaching the global optimum is examined.

VI.3 Recommended Future Research

A number of areas for further study are clearly available. They are:

1. Perform sensitivity analyses for the two algorithms in Chapter III and compare their relative efficiency.

2. Due to the nature of the minisum problem, an efficient global algorithm awaits a breakthrough in non-convex programming. In the meantime, efforts toward developing good bounds using a branch and bound technique would be fruitful.

3. With regard to escaping local optima in the minisum problem, investigate development of existing techniques to capitalize on the structure of the problem.

4. Develop an efficient technique to solve the single facility minimax problem with equally weighted demand points when the demand points form a global subset. Based on comments in Chapter IV, success in this endeavor awaits a breakthrough in solution of the "noxious facility" location problem.

5. Develop a method to solve the weighted single facility and multifacility minimax problem in the great circle metric. It does not appear likely that the geometric approach would be feasible for such problems.

6. Recognizing that in practical applications one is faced with the fact that land covers only a small part

of the earth, develop methods which will consider constraints as to location of the servicing facility (besides the one restricting it to the sphere's surface).

REFERENCES

Angel, Shlomo, and Hyman, Geoffrey M. "Transformations and Geographic Theory," Geographic Analysis, Vol. 4, No. 4, 1972, pp. 350-367.

Bazaraa, Mokhtar S. "An Efficient Cyclic Coordinate Method for Optimizing Penalty Functions," Naval Research Logistics Quarterly, Vol. 22, No. 2, 1975, pp. 399-404.

Blumenthal, Leonard M. "Global Subsets of the Sphere," Archiv der Mathematik, Vol. 7, 1956, pp. 367-373.

_______. A Modern View of Geometry. San Francisco: W. H. Freeman and Co., 1961.

Blumenthal, L. M., and Wahlin, G. E. "On the Spherical Surface of Smallest Radius Enclosing a Bounded Subset of n-Dimensional Space," American Mathematical Society Bulletin, Vol. 47, 1941, pp. 771-777.

Burstall, R. M., Leaver, R. A., and Sussams, J. E. "Evaluation of Transport Costs for Alternate Factory Sites," Operational Research Quarterly, Vol. 13, No. 4, 1962, pp. 345-54.

Chapelle, R. A. "Location of Central Facilities--Heuristic Algorithms for Large Systems," Ph.D. dissertation, University of Oklahoma, 1969.

Chrystal, G. "On the Problem to Construct the Minimum Circle Enclosing n Given Points in a Plane," Proceedings of the Edinburgh Mathematical Society, Vol. 3, 1885, pp. 30-33.

Clough, D. J. "Asymptotic Extreme Value Sampling Theory for Estimation of Global Maximum," Canadian Operations Research Society Journal, Vol. 5, 1969.

Cooper, Leon. "Location-Allocation Problems," Operations Research, Vol. 11, No. 3, 1963, pp. 331-344.

Cooper, Leon. "Heuristic Methods for Location-Allocation Problems," SIAM Review, Vol. 6, 1964, pp. 37-53.

_______. "Solutions of Generalized Locational Equilibrium Models," Journal of Regional Service, Vol. 7, 1967, pp. 9-11.

Cotter, Charles H. The Astronomical and Mathematical Foundations of Geography. New York: American Elsevier Publishing Co., 1966.

Courant, R., and Robbins, H. What Is Mathematics? London: Oxford University Press, 1941.

Curry, G. L., and Skeith, R. W. "A Dynamic Programming Algorithm for Facility Location and Allocation," AIIE Transactions, Vol. 1, No. 2, 1969, pp. 133-138.

Deetz, C. H., and Adams, O. S. Elements of Map Projection. U.S. Department of Commerce, Coast and Geodetic Survey, Special Publication #68. Washington: U.S. Government Printing Office, 1945.

Eilon, S., Watson-Gandy, C.D.T., and Christofides, N. Distribution Management: Mathematical Modeling and Practical Analysis. New York: Hafner Publishing Company, Inc., 1971.

Elzinga, D. J., and Hearn, D. W. "Geometrical Solution for Some Minimax Location Problems," Transportation Science, Vol. 6, No. 4, 1972a, pp. 379-394.

Elzinga, D. J., and Hearn, D. W. "The Minimum Covering Sphere Problem," Management Science, Vol. 19, No. 1, 1972b, pp. 96-104.

Eyster, J. W., White, J. A., and Wierville, W. W. "On Solving Multifacility Location Problems Using a Hyperboloid Approximation Procedure," AIIE Transactions, Vol. 5, No. 1, 1973, pp. 1-6.

Francis, R. L., and Goldstein, J. M. "Location Theory: A Selective Bibliography," Operations Research, Vol. 22, 1974, pp. 400-409.

Francis, R. L., and White, J. A. Facility Layout and Location: An Analytical Approach. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1974.

Gilbert, E. N. "Distortion in Maps," SIAM Review, Vol. 16, No. 1, 1974, pp. 47-62.

Haley, K. B. "The Siting of Depots," The International Journal of Production Research, Vol. 2, No. 1, 1962, pp. 41-45.

Hartman, James K. "Some Experiments in Global Optimization," Naval Research Logistics Quarterly, Vol. 20, No. 3, 1973, pp. 569-576.

Hearn, Donald W. "Minimum Covering Spheres," Ph.D. dissertation, John Hopkins University, 1971.

Henrici, Peter. Elements of Numerical Analysis. New York: Wiley & Sons, Inc., 1964.

Hesse, Rick. "A Heuristic Procedure for Estimating a Global Solution of Non-convex Programming Problems," Operations Research, Vol. 21, No. 6, 1973, pp. 1267-1279.

Hilbert, D., and Cohn-Vossen, S. Geometry and the Imagination. New York: Chelsea Publishing Company, 1952.

Hitchings, G. G. "Analogue Techniques for the Optimal Location of a Main Facility in Relation to Ancillary Facilities," The International Journal of Production Research, Vol. 7, No. 3, 1969, pp. 189-197.

Katz, I. Norman. "On the Convergence of a Numerical Scheme for Solving Some Locational Equilibrium Problems," SIAM Journal of Applied Mathematics, Vol. 17, 1969, pp. 1224-1231.

_______. "Local Convergence in Fermat's Problem," Mathematical Programming, Vol. 6, 1974, pp. 89-104.

Katz, I. N., and Cooper, Leon. "Optimal Location on a Sphere," paper presented at National ORSA Conference, Chicago, 1975.

Katz, I. N., and Cooper, Leon. "Optimum Location on a Sphere," paper presented at National ORSA Conference, Philadelphia, 1976.

Kay, David C. College Geometry. New York: Holt, Rinehart and Winston, 1969.

Kreysig, Erwin. Introduction to Differential Geometry and Riemannian Geometry. Toronto: University of Toronto Press, 1968.

Kuenne, R. E., and Soland, R. M. The Multisource Weber Problem: Exact Solutions by Branch and Bound. IDA Economic Series, Arlington: Institute for Defense Analyses, 1971.

Kuenne, R. E., and Soland, R. M. "Exact and Approximate Solutions to the Multisource Weber Problem," Mathematical Programming, Vol. 3, 1972, pp. 193-209.

Kuhn, H. W. "On a Pair of Dual Nonlinear Programs," in J. Abadie (ed.), Nonlinear Programming, 1967, pp. 37-54.

________. "A Note on Fermat's Problem," Mathematical Programming, Vol. 4, 1973, pp. 98-107.

Kuhn, H. W., and Kuenne, R. E. "An Efficient Algorithm for the Numerical Solution of the Generalized Weber Problem in Spatial Economics," Journal of Regional Science, Vol. 4, 1962, pp. 21-33.

Lea, A. C. "Location-Allocation Systems: An Annotated Bibliography," Discussion Paper No. 13, Department of Geography, Toronto: University of Toronto, 1973.

Liau, T. L. "Confidence Regions for Global Optima in NLP," Texas A & M University, Project THEMIS, Report No. 43, 1973.

Loginov, È. A. "An Algorithm for a Certain Čebyšev Problem," Èkonomika i Mathematiceskie Metody (Russian), Vol. 5, 1969, pp. 113-119).

Love, Robert F. "A Note on the Convexity of the Problem of Siting Depots," The International Journal of Production Research, Vol. 6, No. 2, 1967, pp. 153-154.

Lyusternik, L. A. Shortest Paths Variational Problems. New York: The Macmillan Company, 1964.

Maling, D. H. Coordinate Systems and Map Projections. London: George Philip and Son, Ltd (1973).

Meschkowski, H. Unsolved and Unsolvable Problems in Geometry. New York: Frederick Ungar Publishing Co., 1966.

Milnor, John. "A Problem in Cartography," American Mathematical Monthly, Vol. 76, 1969, pp. 1101-1112.

Pritsker, A. A. B., and Ghare, P. M. "Locating New Facilities with Respect to Existing Facilities," AIIE Transactions, Vol. 2, No. 4, 1970, pp. 290-297.

Rademacher, H., and Toeplitz, O. The Enjoyment of Mathematics. Princeton, N.J.: Princeton University Press, 1957.

Reklaitis, G. V., and Phillips, D. T. "Survey of Nonlinear Programming," AIIE Transactions, Vol. 7, No. 3, 1975, pp. 235-256.

ReVelle, C., Marks, D., and Liebman, J. C. "An Analysis of Private and Public Sector Location Models," Management Science, Vol. 16, 1970, pp. 692-707.

ReVelle, C., and Swain, R. "Central Facilities Location," Geographical Analysis, Vol. 2, No. 1 (1970), pp. 30-42.

Robinson, D. F. "Optimization of Finite Sets," Proceedings of the First Australian Conference on Combinatorial Mathematics. Newcastle: University of Newcastle Research Associates, 1972, pp. 39-52.

Scott, Allen J. "Location-Allocation Systems: A Review," Geographical Analysis, Vol. 2, 1970, pp. 95-119.

Shannon, R. E., and Ignizio, J. P. "A Heuristic Programming Algorithm for Warehouse Location," AIIE Transactions, Vol. 2, No. 4, 1970, pp. 334-339.

Sivazlian, B. D., and Stanfel, L. E. Optimization Techniques in Operations Research. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

Smallwood, R. D. "Minimax Detection Station Placement," Operations Research, Vol. 13, 1965, pp. 632-646.

Sokolowsky, Daniel. "A Note on the Fermat Problem," American Mathematical Monthly, Vol. 83, No. 4, 1976, p. 276.

Sylvester, J. J. "A Question in the Geometry of Situation," Quarterly Journal of Pure and Applied Mathematics, Vol. 1, 1857, p. 79.

_______. "On Poncelet's Approximate Linear Valuation of Surd Forms," Philosophical Magazine, Vol. 20, Series 4, 1860, pp. 203-222.

Tóth, L. Fejes. "Exploring a Planet," American Mathematical Monthly, Vol. 80, No. 9, 1973, pp. 1043-1044.

Vergin, R. C., and Rogers, J. D. "An Algorithm and Computational Procedure for Locating Economic Facilities," Management Science, Vol. 13, No. 6, 1967, pp. 3240-3254.

Warntz, William. "Global Science and the Tyranny of Space," The Regional Science Association Papers, Vol. 19, 1967, pp. 7-19.

Warntz, W., and Wolff, Peter. Breakthroughs in Geography. New York: The New American Library, Inc., 1971.

Weber, Alfred. Über den Standort der Industrien Tübingen, 1909. Translated as Alfred Weber's Theory of the Location of Industries. Chicago: University of Chicago Press, 1929, by C. J. Friedrich.

Weiszfeld, E. "Sur le Point pour Lequel la Somme des Distances de n Points Donnés est Minimum," Tohoku Mathematics Journal, Vol. 43, 1936, pp. 355-386.

Wendell, Richard E. "Some Aspects in the Theory of Location," Ph.D. dissertation, Northwestern University, 1971.

Wendell, Richard E., and Hurter, Arthur P., Jr. "Location Theory, Dominance, and Convexity," Operations Research, Vol. 21, No. 1, 1973, pp. 314-320.

Zangwill, Willard I. Nonlinear Programming: A Unified Approach. Englewood Clifss, N.J.: Prentice-Hall, Inc., 1974.

APPENDIX A

CONVEXITY THEORY IN SPHERICAL GEOMETRY

A.1 Preliminary Definitions

In spherical geometry one has the axiomatic framework as set forth by Kay (1969) with the canonical model being the unit sphere S^2 in Euclidean 3-space given by $|x|_{\ell_2} = 1$. In this model, the spherical metric $\rho(x,y)$ for each x and y in S^2 is the length of the minor arc of the great circle joining x and y, and α, the least upper bound for the metric, has the value π. Further, the lines of the geometry are great circles, and if two points x and y are not antipodal $[\rho(x,y) \neq \pi]$ the line through x and y is unique; in this case the line is denoted $\overleftrightarrow{xy}$. Each line can be coordinatized from the set of numbers $\{\lambda | -\pi < \lambda \leq \pi\}$ (the coordinate set), with any point as origin (zero coordinate) and such that if the coordinates of u and v are λ and μ, then $\rho(u,v) = |\lambda-\mu|$ if $|\lambda-\mu| \leq \pi$, or $\rho(u,v) = 2\pi - |\lambda-\mu|$ if $|\lambda-\mu| > \pi$. Betweenness can be defined by asserting that z is between x and y, if and only if x, y, and z are distinct points and $\rho(x,z) + \rho(z,y) = \rho(x,y)$. If z is between x and y we denote the relationship by (xzy).

For any two points x and y such that $0 < \rho(x,y) < \pi$ the segment joining x and y, denoted $\overline{xy}$, is the set of points $\{z | z=x,\ z=y, \text{ or } (xzy)\}$.

Definition A.1 A set C in S^2 is said to be convex if for each two points x and y in C such that $0 < \rho(x,y) < \pi$ then the segment $\overline{xy} \subset C$. The convex hull of a set A in S^2 is the intersection of all convex sets containing A, denoted conv A.

It is obvious that an arbitrary intersection of convex sets is convex; hence conv A is a convex set for each subset A of S^2. A set is convex iff it equals its own convex hull. Note that the above definition of convexity allows the following examples to be convex sets: A hemisphere, a great circle, a lune, and any pair of antipodal points. The convex hull of 2 points is either a segment or the two points themselves; the convex hull of a closed hemisphere and a point not on it is the entire sphere S^2.

One of the important axioms concerning betweenness on S^2 is the so-called plane separation property: Each line L in S^2 divides the points of S^2 into three pairwise disjoint convex sets L, H_1, and H_2 such that if $x \in H_1$ and $y \in H_2$ there exists a point $z \in L$ such that (xzy). The sets H_1, H_2 are referred to as the half-spaces determined by L, and two points which lie in the same half-space determined by L are said to lie on the same side of L. Note in passing that a half-space is a hemisphere on S^2 with its boundary (great circle) deleted.

A <u>half-line</u> or <u>ray</u> is in a subset L' of a line L of the form $\{u[\lambda] | \lambda \geq 0\}$ where $u[\lambda]$ denotes a coordinate system for L; a half-line always consists of a (closed) semi-great circle. The point $u[0] = x$ is called the <u>origin</u> of the ray, and if y is a point on L' such that $0 < \rho(x,y) < \pi$, the ray is denoted $\overrightarrow{xy}$. It is shown by Kay (1969) that if $z \varepsilon \overrightarrow{xy}$ with $0 < \rho(x,z) < \pi$, then $\overrightarrow{xy} = \overrightarrow{xz}$.

An <u>angle</u> is the union of two rays having the same origin, the two rays being called its <u>sides</u> and the common origin its <u>vertex</u>. If $\overrightarrow{xy}$ and $\overrightarrow{xz}$ are the sides of an angle the notation $\overline{\sphericalangle yxz}$ is used. The measure $\sphericalangle yxz$ of the angle $\overline{\sphericalangle yxz}$ is the Euclidean measure of the angle formed by the tangents to the sides $\overrightarrow{xy}$ and $\overrightarrow{xz}$ at x in 3-space. Note that a straight angle (having measure π) is a line on S^2, and that every other angle is the boundary of a region on S^2 usually called a <u>lune</u> or spherical <u>section</u>.

Given three distinct rays $\overrightarrow{xu}$, $\overrightarrow{xv}$, $\overrightarrow{xw}$ having the same origin x, we say that $\overrightarrow{xv}$ lies between $\overrightarrow{xu}$ and $\overrightarrow{xw}$, and we write $(\overrightarrow{xu}\ \overrightarrow{xv}\ \overrightarrow{xw})$, if $\sphericalangle uxv + \sphericalangle vxw = \sphericalangle uxw$. The <u>interior</u> of an angle $\overline{\sphericalangle yxz}$ is the set of all points lying on rays which lie between the sides $\overrightarrow{xy}$ and $\overrightarrow{xz}$. (See Kay (1969) for betweenness properties of segments, rays, and triangles.)

If x, y, z are 3 distinct noncollinear points of S^2 it follows that $0 < \rho(x,y) < \pi$, $0 < \rho(x,z) < \pi$, and

$0 < \rho(y,z) < \pi$. Thus, the segments $\overline{xy}$, $\overline{yz}$, $\overline{zx}$ exist, their union defining a set called a <u>triangle</u>, denoted Δxyz. If for each point x and line L not through it $H(x,L)$ denotes the half-plane determined by L and containing x, the <u>interior</u> of triangle xyz is by definition the set $H(x,\overleftrightarrow{yz}) \cap H(y,\overleftrightarrow{xz}) \cap H(z,\overleftrightarrow{xy})$. It can be shown that p is an interior point of triangle xyz iff the line $\overleftrightarrow{xp}$ intersects $\overline{yz}$ at a point w such that (xpw) and (ywz), and similarly with lines $\overleftrightarrow{yp}$ and $\overleftrightarrow{zp}$. The Crossbar Principle (Kay 1969) then shows that $(\overrightarrow{xy}\ \overrightarrow{xp}\ \overrightarrow{xz})$.

Finally, a <u>circle</u> on S^2 is the set of all points at a fixed distance (<u>radius</u>) from a fixed point (<u>center</u>). Note that a circle with radius $\pi/2$ is also a line on S^2; the center of any other circle may be chosen in such a manner that the radius is less than $\pi/2$. The <u>interior</u> of a circle of radius r is the set of all points at a distance less than r from the center. A <u>disk</u> on S^2 is a circle of radius $< \pi/2$ and its interior; the <u>center</u> of the disk is the center of the circle defining it. (Circles and their interiors are not convex sets if their radii are greater than $\pi/2$.)

A.2 <u>Some Topology on S^2</u>

Given any set A on S^2 a point x is called an interior point of A if x is the center of some disk which is entirely contained by A. A point x is called an <u>exterior point</u> of A if x is the center of some disk which is entirely devoid of points of A. A point which is neither an interior or exterior point of A is called a boundary point. We denote by int A, ext A, and bd A, respectively, the set of all interior points of A, the set of all exterior points of A, and the set of all boundary points of A, called the interior, exterior, and boundary of A. A set K is closed if bd $K \subset K$ (K contains its boundary points). A set is open if int $K \supset K$ (every point of K is an interior point of K). It easily follows that for any set K in S^2, int $K = K \setminus$ bd $K \equiv$ the set of all points of K which do not belong to bd K.

A set K in any metric space is called compact if for each sequence of points $\{x_n\}$ in K some subsequence $\{x_{n_k}\}$ converges in the metric to a point x in K (that is, $\lim_{k\to\infty} \rho(x,x_{n_k}) = 0$). Since S^2 is itself a compact subset of a metric space it follows that any closed subset K of S^2 is compact. To see this, assume $\{x_n\} \subset K$ and that $\lim_{k\to\infty} \rho(x,x_{n_k}) = 0$ for some $x \in S^2$ and some subsequence $\{x_{n_k}\}$ of $\{x_n\}$. Then if $x \notin K$ certainly x is a boundary point of K: every disk with x as center contains points in K (namely, the points x_{n_k} for large enough k such that $\rho(x,x_{n_k}) <$ radius of the disk), and a point not in K (namely, x itself). Since K is closed, bd $K \subset K$ or $x \in K$, a contradiction. Hence, $x \in K$ and $\{x_{n_k}\}$ converges to a point in K proving that K is compact.

A well-known result of topology is that if f is any real-valued continuous function defined on a compact subset K of a metric space, there exist points k_1 and k_2 in K such that $f(k_1)$ is the absolute minimum of f on K and $f(k_2)$ is the absolute maximum.

<u>Definition A.2</u> A (topological) <u>separation</u> of a subset A of a metric space is a pair of disjoint open sets U, V such that U and V both intersect A and $A \subset U \cup V$. A set A is <u>connected</u> if it has no such separation. (Thus, A is connected if and only if for each pair of disjoint open sets U, V such that $A \subset U \cup V$, then either $A \subset U$ or $A \subset V$.)

Common examples of connected sets in R^2 are points, segments, arcs, continuous curves, graphs of continuous functions, and convex sets. To see that a segment is connected, let its endpoints be p and q, and let the segment be parametrized by $\{\lambda \mid 0 \leq \lambda \leq \text{length of } \overline{pq} \equiv \beta\}$, with $p_\lambda \in \overline{pq}$ such that $p_0 = p$ and $p_\beta = q$. If (U,V) is a separation of $\overline{pq}$ then without loss of generality, one can assume $p_{\lambda_1} \in U$, $p_{\lambda_2} \in V$, with $\lambda_1 < \lambda_2$. Let λ_0 be the least upper bound of the set $\lambda \leq \lambda_2$ such that $p_\lambda \in U$. Then $p_{\lambda_0} \notin V$, or else there exists a disk D centered at p_{λ_0} contained in V; by definition of λ_0 a sequence $\{\lambda_n\} \to \lambda_0$ exists with $p_{\lambda_n} \in U$, and hence $\lim_{n\to\infty} p_{\lambda_n} = p_{\lambda_0}$ so $p_{\lambda_n} \in D \subset V$ for all sufficiently large n, contradicting the disjointness of U and V. Hence $p_{\lambda_0} \in U$. Since $\lambda_0 \leq \lambda_2$ and $p_{\lambda_2} \in V$ one must have $\lambda_0 \neq \lambda_2$, and $p_\lambda \in V$ for $\lambda_0 < \lambda \leq \lambda_2$. Let $\{\lambda_n'\}$ be a sequence converging to λ_0 such that $\lambda_0 < \lambda_n' < \lambda_2$. Then $\lim_{n\to\infty} p_{\lambda_n'} = p_{\lambda_0}$

and in the same manner that $p_{\lambda_n} \in V$ was proved one obtains the contradiction $p_{\lambda'_n} \in U$. Therefore, no separation of $\overline{pq}$ exists and $\overline{pq}$ is connected. (The definition of convexity can now be shown to imply the connectedness of any convex set.)

The type argument used above to prove the connectedness of a line segment in R^2 will occur later in a slightly different context. First an obvious property of connected sets is established.

Lemma A.1 If A is any connected set in R^2 which contains two points on opposite sides of a line L, then L passes through a point of A.

Proof: The two sides of L form two open sets U and V such that $U \cup V \cup L = R^2$. Hence $A \subset U \cup V \cup L$. If $A \cap L = \phi$ then $A \subset U \cup V$, with A meeting both U and V. Hence (U,V) separates A, denying the connectedness of A.

Recall that if f is any real valued function of a real variable, the graph of f, denoted graph f, is the subset of R^2 consisting of all points with coordinates $(\xi, f(\xi))$, where ξ is real and varies on the domain of f. It will be convenient to augment this set in the case of a discontinuous function.

Definition A.3 Let f be any bounded real-valued function of a real variable defined on an interval $\alpha \leq \xi \leq \beta$. The augmented graph of f is the set graph $f \cup T$, denoted by graph*f, where T consists of all line segments joining $(\gamma, f(\gamma))$ with (γ, γ_1) and (γ, γ_2), where $\alpha \leq \gamma \leq \beta$ and $\gamma_1 = \lim\sup_{\xi\to\gamma} f(\xi)$, $\gamma_2 = \lim\inf_{\xi\to\gamma} f(\xi)$. (Recall that $\lim\sup_{\xi\to\gamma} f(\xi)$ and $\lim\inf_{\xi\to\gamma} f(\xi)$ are respectively the maximal and minimal values to which sequences $f(\xi_n)$ converge for $\{\xi_n\}$ converging to γ.) Let T_γ denote the segment (perhaps degenerate) joining the three collinear points $(\gamma, f(\gamma))$, (γ, γ_1) and (γ, γ_2) for each $\alpha \leq \gamma \leq \beta$.

Lemma A.2 The augmented graph in R^2 (graph*f) of any bounded real-valued function f of a real variable is connected.

Proof: Using the same notation as above, for each $(\gamma, \delta) \in$ graph*f then $\alpha \leq \gamma \leq \beta$ and (γ, δ) lies on T_γ. (If f is continuous at γ, T_γ = the single point $(\gamma, f(\gamma))$.) Suppose graph*f is not connected. Then graph*f has a separation (U,V) with U,V disjoint open subsets of R^2, graph*f $\subset U \cup V$, and graph*f meets both U and V. Since T_γ is connected for all γ then either $T_\gamma \subset U$ or $T_\gamma \subset V$ for all $\alpha \leq \gamma \leq \beta$ Since all points of graph*f lie on some T_γ then for certain values γ_1 and γ_2, $T_{\gamma_1} \subset U$ and $T_{\gamma_2} \subset V$; one can assume that $\gamma_1 < \gamma_2$. Let γ_0 be the least upper bound of all $\gamma \leq \gamma_2$ such that $T_\gamma \subset U$. Then $T_{\gamma_0} \subset U$, for otherwise, $T_{\gamma_0} \subset V$, and by definition of γ_0 there exists a sequence $\{\gamma_n\}$ converging to γ_0 such that $\gamma_n \leq \gamma_0$ and $T_{\gamma_n} \subset U$. Let $(\gamma_n, \delta_n) \in T_{\gamma_n}$; since $\{\delta_n\}$ is bounded there exists a subsequence $\{\delta_k\}$ of $\{\delta_n\}$ (writing k in place of n_k)

such that $\{\delta_k\}$ converges to some δ_0. Hence $\liminf_{\xi\to\gamma_0} f(\xi) \leq \delta_0 \leq \limsup_{\xi\to\gamma_0} f(\xi)$ and $(\gamma_0,\delta_0) \in T_{\gamma_0}$. But if $T_{\gamma_0} \subset V$ there is a disk D in R^2 entered at (γ_0,δ_0) and contained in V. Since $\lim_{k\to\infty} (\gamma_k,\delta_k) = (\gamma_0,\delta_0)$ one has $(\gamma_k,\delta_k) \in D \subset V$ for all sufficiently large k. But $(\gamma_k,\delta_k) \in T_{\gamma_k} \subset U$, contradicting the disjointness of U and V. Therefore, $T_{\gamma_0} \subset U$ and since $T_{\gamma_2} \subset V$, $\gamma_0 \neq \gamma_2$, or $\gamma_1 \leq \gamma_0 < \gamma_2$, and for all $\gamma_0 < \gamma \leq \gamma_2$, $T_\gamma \subset V$ by definition of γ_0. Now let $\{\gamma_n'\}$ be a sequence converging to γ_0 such that $\gamma_n' > \gamma_0$ for all n. Thus, $T_{\gamma_n'} \subset V$. Let $(\gamma_n',\delta_n') \in T_{\gamma_n}$; in the same manner as before it can be shown that for a certain subsequence $\{(\gamma_k',\delta_k')\}$ of $\{(\gamma_n',\delta_n')\}$ the point (γ_k',δ_k') belongs to both U and V for all k sufficiently large, in contradiction. This final contradiction proves that there is no separation of graph*f, and graph*f is connected.

Observe that if f is continuous, graph*f = graph f, and one obtains the well-known result:

<u>Corollary A.2.1</u> The graph in R^2 of a continuous real function is connected.

A.3 <u>Support Lines</u>

Suppose C is a convex set on S^2 and L is a line such that C lies entirely on L or on one side of L (that is, $C \subset L \cup H_1$, where H_1 is one of the half-spaces determined by L) and L intersects C at least at one point. Then L is called a <u>line of support</u> for C. Two fundamental lemmas concerning support lines are now stated and proved.

Lemma A.3 Given any convex closed set $C \subset S^2$ and a point p in its boundary, there exists a line of support for C passing through p.

Proof: If C is zero- or one-dimensional the claim is obvious since $C \subset L$ for some line L. Hence assume C is two-dimensional and $p \in$ bd C. Choose any line L_0 through p. If L_0 is not already a line of support there exist two points of C on opposite sides of L_0, q and s, not on L_0, and thus a point r on L_0 such that (qrs); by convexity of C, $r \in C$. Thus, the set of rays $\overrightarrow{px}$ for which $x \in C \cap H$ where H is the opposite side of L_0 as s, is nonempty. Accordingly, set

$$\theta_0 = \sup\{\theta = \measuredangle xpr\}, \qquad 0 < \theta_0 \leq \pi,$$

for $x \in C \cap H$. It is claimed that if ray $\overrightarrow{px_0}$ is determined in $L_0 \cup H$ such that $\measuredangle x_0pr = \theta_0$ then line $\overleftrightarrow{px_0} \equiv L$ is a line of support of C through p.

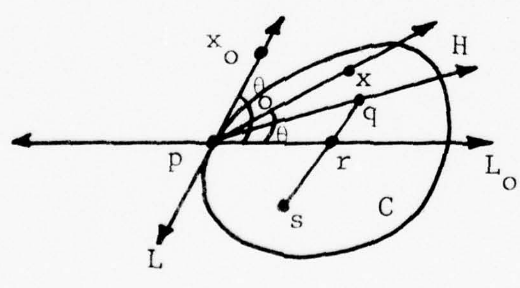


Figure A.1.

If L is not a line of support there exist points of C on opposite sides of L, and thus a point $y \in C$ lies on the r'-side of L where ray $\overrightarrow{pr'}$ opposes $\overrightarrow{pr}$ on L_0. If $y \in H$ then

$(\overrightarrow{pr'}\ \overrightarrow{py}\ \overrightarrow{px_0})$ follows from properties of betweenness (Kay 1969) and $\theta = \measuredangle ypr > \measuredangle x_0pr = \theta_0$, contradicting $\theta_0 \geq \theta$

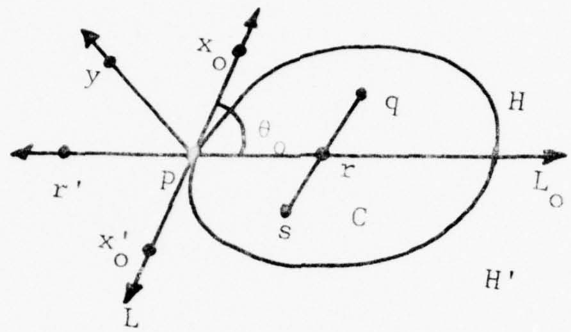


Figure A.2.

(definition of θ_0). Hence $y \in L_0 \cup H'$ where H' is the x_0'-side of L_0. But $y \in L_0$ implies p is interior to Δyqs and

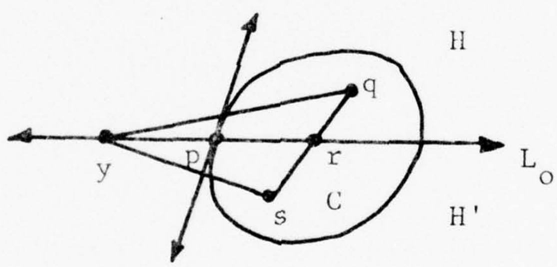


Figure A.3.

since C is convex with y, q, and s in C, Δyqs and its interior belong to C. Hence, $p \in \text{int } C$, a contradiction (an interior point cannot be a boundary point). Hence $y \in H'$ and the ray $\overrightarrow{py'}$ opposing $\overrightarrow{py}$ lies on the H side of L_0. Since (ypx') holds for some point $x' \in C$ then $x' \in C \cap H$ and $\theta_0 > \theta' = \measuredangle x'pr$. By definition of θ_0 there exists $x \in C \cap H$ such that $\theta' < \theta < \theta_0$ where $\theta = \measuredangle xpr$. Then it follows that (Kay 1969, p. 63) $(\overrightarrow{px}\ \overrightarrow{pr'}\ \overrightarrow{py})$ and that $p \in \text{int } C$, a contradiction. Therefore, L is a line of support through p.

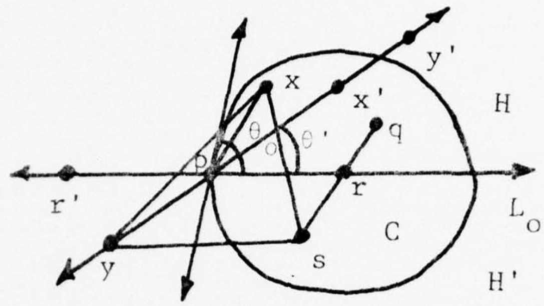


Figure A.4.

<u>Lemma A.4</u> Let K be a closed subset of S^2, with $x \notin K$. There exists $p \in \text{bd}\, K$ such that the perpendicular L to line $\overleftrightarrow{xp}$ at p is a line of support of K.

<u>Proof</u>: Define the function f on K by setting $f(k) = \rho(x,k)$, for each $k \in K$. It is well known that the metric itself is continuous in its own topology (that defined by disks), so it follows that f is a continuous, real-valued function defined on a compact set. By an earlier comment (compactness, Section A.2) f takes on its absolute minimum at some point $p \in K$. Since $\rho(p,x) \leq \rho(k,x)$ for all $k \in K$ it follows that p is not interior to K. Hence $p \in \text{bd}\, K$. Consider the perpendicular L to $\overleftrightarrow{xp}$ at p. If L were not a line of support there would exist a point $y \in K$ on the x-side of L ($y \notin L$), and hence (Figure A.5) $\measuredangle xpy < \pi/2$. Now, $x \notin \overleftrightarrow{py}$, for otherwise $y \in \overrightarrow{xp}$ which means (since $x \neq y$) either (yxp) or $x \in \overline{py} \subset K$ by convexity

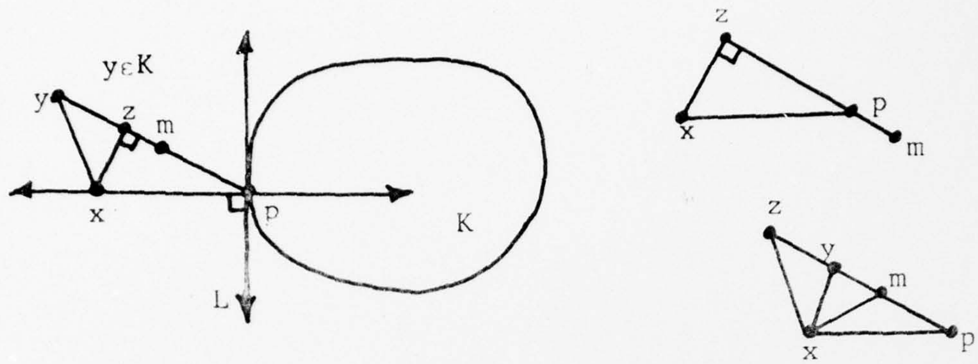


Figure A.5.

of K, a contradiction, or (xyp) which would imply $\rho(x,y) < \rho(x,p)$, a contradiction. Take first the case $\rho(x,p) < \pi/2$. Locate z, the foot of x, on line $\overleftrightarrow{py}$. Then since $\rho(x,z) \leq \rho(x,p) < \pi/2$ one has $\rho(x,z) < \pi/2$ and it follows (Kay 1969, p. 114) that $\rho(x,z) < \rho(x,p)$. Hence $z \notin K$ or else the definition of $p \in K$ is contradicted. Thus $y \neq z$ and $\rho(x,z) < \rho(x,y)$. Let m be the midpoint of segment $\overline{py}$; then $\rho(p,m) = \rho(m,y) < \pi/2$ since $\rho(p,y) = \pi$ implies $y \in \overleftrightarrow{px}$. By Kay (1969, p. 52) either (zpm), $z = p$, (mzp), $m = z$, (mzy), $z = y$, or (myz) holds. The cases $z = p$ and $z = y$ are ruled out since $z \notin K$. Also, since $p, y \in K$ and K is convex one cannot have $z \in \overline{py}$; therefore (mzp), $m = z$, and (mzy) cannot hold. This leaves (zpm) or (myz). If (zpm) then since the sides of triangle xpz are each $< \pi/2$ the Exterior Angle Theorem (Kay 1969) gives $\measuredangle mpx > \measuredangle xzp = \pi/2$. But $\measuredangle mpx = \measuredangle ypx < \pi/2$ (since y and x are on the same side of L), which is a contradiction. This leaves the case (myz). Here, $\rho(x,z) < \pi/2$ and $\rho(x,p) < \pi/2$ imply $\rho(x,m) < \pi/2$ and, in turn, $\rho(x,y) < \pi/2$ by Kay (1969, p. 98). Therefore, the Exterior Angle Theorem produces $\measuredangle xyp = \measuredangle xym > \measuredangle xzy = \pi/2 > \measuredangle xpy$ and by Theorem 31.5 in Kay (1969), $\rho(x,y) < \rho(x,p)$, a contradiction since $y \in K$. This completes the case $\rho(x,p) < \pi/2$. If $\rho(x,p) \geq \pi/2$ then the situation is as pictured in the figure below.

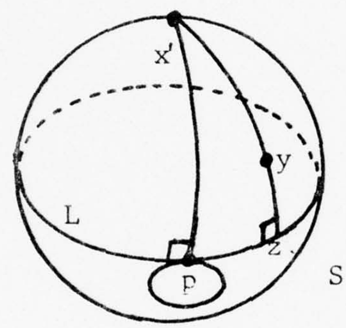


Figure A.6.

Here, choose x' on $\overleftrightarrow{px}$ with $\rho(x',p) = \pi/2$ and z on L such that (x'yz). Then $\rho(x,y) < \rho(x,z) + \rho(x',x) \leq \rho(x,p)$, a contradiction. This is seen in the following:

$$\begin{aligned}\rho(x,y) &\leq \rho(x,x') + \rho(x',y)\\ &< \rho(x,x') + \rho(x',z)\\ &= \rho(x,x') + \rho(x',p)\\ &= \rho(x,p)\end{aligned}$$

Therefore, no such point $y \in K$ exists and L is a line of support.

<u>Lemma A.5</u> Given a closed convex set K of S^2 consisting of more than one point and a point $x \notin K$ such that no line through x contains K, there exist exactly two

distinct rays $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$ with $b_k \in \text{bd } K$, $k = 1,2$, such that the lines $\overleftrightarrow{xb_k}$ are support lines, and the ray $\overrightarrow{xu}$ for each $u \in K$ either coincides with $\overrightarrow{xb_1}$, or $\overrightarrow{xb_2}$, or lies between $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$.

Proof: Since no line through K contains x and K is not a point, there is a segment $\overline{y_1y_2} \subset K$ such that $z \in \overline{y_1y_2}$, and y_1 and y_2 lie on opposite sides of line $\overleftrightarrow{xz}$. Define the bounded nonempty sets:

$$S_k = \{\theta_k = \measuredangle zxu_k \mid u_k \in K,\ u_k \text{ lies on the } y_k\text{-side of } \overleftrightarrow{xz},\ k = 1,2\}$$

Let $\theta_k^* < \pi$ be the least upper bound of S_k, $k = 1,2$. Choosing a sequence $\{b_n^k\}$ such that $b_n^k \in \overrightarrow{xu_k} \cap K$ for each k, a subsequence $\{b_m^k\}$ converges to some point $b_k \in K$ (since K is compact) such that $\theta_k^* = \measuredangle zxb_k$, $k = 1,2$.

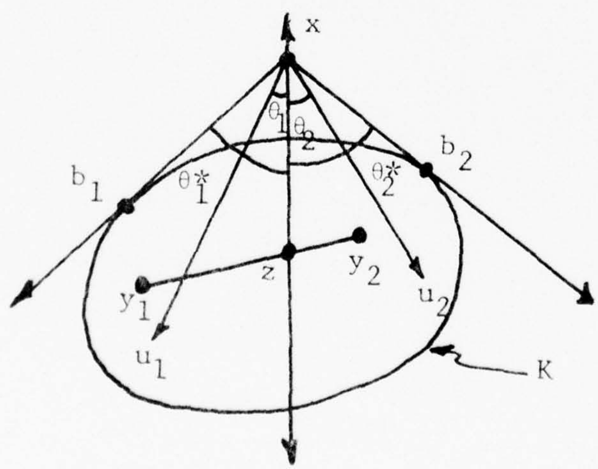


Figure A.7.

Observe that $b_k \in \text{bd } K$, for, if $b_k \notin \text{bd } K$ then b_k is an interior point of K, in which case a disk D about b_k

lies in K; but then there exist $u \in D \cap V$ such that ray $\overrightarrow{xb_k}$ lies between $\overrightarrow{xu}$ and $\overrightarrow{xz}$ and u is on the y_k-side of $\overleftrightarrow{xz}$ with $\theta = \measuredangle zxu = \measuredangle zxb_k + \measuredangle b_kxu > \measuredangle zxb_k = \theta_k^*$, contradicting the fact that θ_k^* is an upper bound for S_k.

Now, if $\overleftrightarrow{xb_k}$ is not a support line of K there exists points u_1, u_2 of K on opposite sides of line $\overleftrightarrow{xb_k}$, and the segment $\overline{u_1u_2}$ intersects $\overleftrightarrow{xb_k}$ at a point u_3; if $u_3 \notin$ ray $\overrightarrow{xb_k}$ or $u_3 = x$ then $x \in \text{conv}\,\{u_1,u_2,z\} \subset K$ contradicting $x \notin K$. Hence $u_3 \in \overrightarrow{xb_k}$ and $u_3 \neq x$, so that a point $u \in \overline{u_1u_2} \cap K$ such that ray $\overrightarrow{xb_k}$ lies between $\overrightarrow{xu}$ and $\overrightarrow{xz}$ and u is on the y_k side of $\overleftrightarrow{xz}$ with $\theta = \measuredangle zxu > \measuredangle zxb_k = \theta_k^*$ (as before), in contradiction. Therefore, $\overleftrightarrow{xb_k}$ is a support line of K.

Finally, if u is any point of K then u lies both on the b_2 side of $\overleftrightarrow{xb_1}$, and the b_1 side of $\overleftrightarrow{xb_2}$. That is, $u \in [H(\overleftrightarrow{b_1x},b_2) \cap H(\overleftrightarrow{b_2x},b_1)] \cup \overrightarrow{xb_1} \cup \overrightarrow{xb_2}$ and u lies on $\overrightarrow{xb_1}$, $\overrightarrow{xb_2}$, or is an interior point of the angle b_1xb_2. Hence either $\overrightarrow{xu}$ coincides with $\overrightarrow{xb_1}$ or $\overrightarrow{xb_2}$, or $\overrightarrow{xu}$ lies between $\overrightarrow{xb_1}$ and $\overrightarrow{xb_2}$.

(Note that, in particular, $z \in K$ so $(\overrightarrow{xb_1}\ \overrightarrow{xz}\ \overrightarrow{xb_2})$ or $\measuredangle zxb_1 + \measuredangle zxb_2 = \measuredangle b_1xb_2 \leq \pi$ and $\theta_1^* + \theta_2^* \leq \pi$.)

<u>Lemma A.6</u> Given a closed convex set V of S^2, with points $x \notin V$ and $b \in \text{bd}\,V$ such that $\overleftrightarrow{xb}$ is a support line to V at b. There exists a point x' on bd V and $u \in \overline{xb}$ such that $\overleftrightarrow{ux'}$ is a support line of V at x' and $\rho(u,x') = \rho(x,u)$.

Proof: Let g be defined as follows: For each $u \in \overline{xb}$ let u' denote that point on $L_u \cap$ bd V nearest u, where L_u denotes the unique line of support of V from u different from the support line $\overleftrightarrow{xb}$ (Lemma A.5). Then take $g(\lambda) \equiv g(u_\lambda) = \rho(u_\lambda, u'_\lambda) - \rho(x, u_\lambda)$, $0 \leq \lambda \leq \beta$, where u_λ denotes a parametrization of segment $\overline{xb}$ such that $\rho(b, u_\lambda) = \lambda$, with $u_0 = b$ and $u_\beta = x$. *Note that*

$g(0) = \rho(b,b') - \rho(x,b) = -\rho(x,b) < 0$, $g(\beta) = \rho(x,x') - \rho(x,x)$
$= \rho(x,x') > 0$.

The augmented graph of g in R^2 (coordinated by (ξ,η)) is connected by Lemma A.2, with the points $(0,g(0))$ and $(\beta,g(\beta))$ on opposite sides of the line $\eta = 0$ (the ξ-axis); hence $(\gamma,0) \in$ graph*g for some γ, $0 \leq \gamma \leq \beta$. If g is continuous at γ then $\liminf_{\lambda\to\gamma} g(\lambda) = g(\gamma) = \limsup_{\lambda\to\gamma} g(\lambda)$ and $(\gamma,0) = (\gamma,g(\gamma))$. In this case one has $g(\gamma) = 0$, or, $\rho(u_\gamma, u'_\gamma) = \rho(x, u_\gamma)$.

However, g may be discontinuous at γ; so it is desired to show that g has a simple jump discontinuity at any such point. From the nature of the boundary of a convex set it is clear that for points u_λ with $\lambda < \gamma$ the corresponding point u'_λ on bd V precedes or coincides with u'_γ, taking a counterclockwise orientation (parametrization) of bd V. For, if u'_λ follows u'_γ on bd V (Figure A.8) then $\overleftrightarrow{bu'_\lambda}$ meets line $\overleftrightarrow{u_\gamma u'_\gamma}$ at a point v such that either $v = u'_\lambda$ or u'_λ is between v and b, since $\overleftrightarrow{u_\gamma u'_\gamma}$ is a line of support of V and the points b and u'_λ cannot lie on opposite sides of $\overleftrightarrow{u_\gamma u'_\gamma}$. By the Crossbar Principle (Kay 1969) and related ideas, the segment $\overline{u_\lambda u'_\lambda}$ meets $\overline{u'_\lambda b}$ at some point w interior to both segments, and hence V contains points u'_γ and b on opposite sides of line $\overleftrightarrow{u_\lambda u'_\lambda}$, contradicting the fact that $\overleftrightarrow{u_\lambda u'_\lambda}$ is a line of support of V.

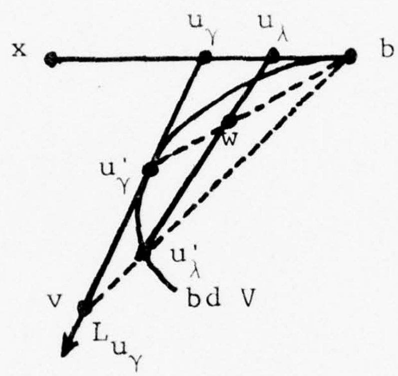


Figure A.8.

Now let γ be a point of discontinuity of g. Let $\lambda \to \gamma^-$ (that is, λ approaches γ from values less than γ). Since bd V is compact some sequence $\{\lambda_n\}$ of λ's is such that $\{u'_{\lambda_n}\}$ converges to a point v_1 bd V. If $\lim\limits_{\lambda\to\infty} u'_\lambda \neq v_1$ then another sequence $\{\mu_n\}$ of λ's is such that $\{u'_{\mu_n}\}$ converges to $v_2 \neq v_1$ on bd V. But the line $\overleftrightarrow{u_{\lambda_n} u'_{\lambda_n}}$ converges to the line $\overleftrightarrow{u_\gamma v_1}$, and this line must be a line of support to V from u_γ. Since there are exactly two such support lines (Lemma A.5), $\overleftrightarrow{u_\gamma v_1}$ must be either L_{u_γ} or $\overleftrightarrow{xb}$; it is obvious that $\overleftrightarrow{u_\gamma v_1} = L_{u_\gamma}$. Similarly, $\overleftrightarrow{u_\gamma v_2} = L_{u_\gamma}$ and hence both v_1 and v_2 lie on $L_{u_\gamma} = \overleftrightarrow{u_\gamma u'_\gamma}$ with v_1 and v_2 coinciding with or preceding u'_γ on bd V. But by definition of u'_γ,

$$\rho(u_\gamma, v_k) \geq \rho(u_\gamma, u'_\gamma), \qquad k = 1,2$$

and if inequality prevails then v_k would follow u'_γ on bd V, which is impossible.

Hence

$$\rho(u_\gamma, v_k) = \rho(u_\gamma, u'_\gamma)$$

with $v_k \in L_{u_\gamma}$. Hence $v_k = u'_\gamma$ for $k = 1,2$. This proves that $\lim_{\lambda\to\gamma^-} u'_\lambda = u'_\gamma$ and hence

$\lim_{\lambda\to\gamma^-} g(\lambda) = \lim_{\lambda\to\gamma^-} [\rho(u_\lambda, u'_\lambda) - \rho(x, u_\lambda)] = \rho(u_\gamma, u'_\gamma) - \rho(x, u_\gamma) = g(\gamma)$.

By an argument exactly similar it can be shown that if u''_γ is that point on $L_{u_\gamma} \cap$ bd V farthest from u_γ then

$$\lim_{\lambda\to\gamma^+} u'_\lambda = u''_\gamma$$

and

$$\lim_{\lambda\to\gamma^+} g(\lambda) = \rho(u_\gamma, u''_\gamma) - \rho(x, u_\gamma) \equiv \alpha_\gamma .$$

Since $\rho(u_\gamma, u''_\gamma) \geq \rho(u_\gamma, u'_\gamma)$ we have

$$\rho(u_\gamma, u''_\gamma) - \rho(x, u_\gamma) \geq g(\gamma)$$

or

$$\lim_{\lambda\to\gamma^+} g(\lambda) = \alpha_\gamma \geq g(\gamma).$$

Hence if g is discontinuous at γ then

$$\lim_{\lambda\to\gamma^+} g(\lambda) = \alpha_\gamma > g(\gamma) = \lim_{\lambda\to\gamma^-} g(\lambda).$$

Thus it follows that

$$\liminf_{\lambda\to\gamma} g(\lambda) = \min\{g(\gamma), \alpha_\gamma\} = g(\gamma)$$

$$\limsup_{\lambda\to\gamma} g(\lambda) = \max\{g(\gamma), \alpha_\gamma\} = \alpha_\gamma$$

and $(\gamma, 0) \in$ graph*g implies $(\gamma, 0)$ lies on the segment determined by the points $(\gamma, g(\gamma))$, (γ, α_γ) and $g(\gamma) \leq 0 \leq \alpha_\gamma$. That is,

$$\rho(u_\gamma, u'_\gamma) - \rho(x, u_\gamma) \leq 0 \leq \rho(u_\gamma, u''_\gamma) - \rho(x, u_\gamma)$$

$$\rho(u_\gamma, u'_\gamma) \le \rho(x, u_\gamma) \le \rho(u_\gamma, u''_\gamma).$$

Since $\rho(u_\gamma, v)$ varies continuously from $\rho(u_\gamma, u'_\gamma)$ to $\rho(u_\gamma, u''_\gamma)$ as v varies on the segment $\overline{u'_\gamma u''_\gamma}$, the ordinary Intermediate Value Theorem guarantees a point v on $\overline{u'_\gamma u''_\gamma} \subset$ bd V such that

$$\rho(u_\gamma, v) = \rho(x, u_\gamma).$$

Thus in all cases it has been proven that there exists a point x' on bd V and u such that $\overleftrightarrow{ux'}$ is a support line of V at x' and $\rho(u, x') = \rho(x, u)$.

<u>Definition A.4</u> The <u>Carathéodory number c</u> is the least integer for which it is true that for any set A and any point $x \in$ conv A there exist c points $a_1, \ldots, a_c$ of A such that $x \in \text{conv}\{a_1, \ldots, a_c\}$.

<u>Lemma A.7</u> The number c (Carathéodory) is equal to 3 in spherical space, as it is on the plane.

<u>Proof</u>: It is clear that $c \ge 3$ by considering the case $A = \{a_1, a_2, a_3\}$, the vertices of a nondegenerate spherical triangle. Hence it suffices to show that $c \le 3$. Let $A \subset S^2$ and suppose $x \in$ conv A; also, suppose $x \notin A$, for otherwise one may simply take $a_1 = a_2 = a_3 = x$ and the result is trivial. Assume first that A is contained by some closed hemisphere. Consider the two cases: $x \in$ bd conv A and $x \in$ int conv A. If $x \in$ bd conv A let L be a line of support of conv A at x and H the A-side of L if $A \not\subset L$; here it follows that $x \in \text{conv}\{a_1, a_2\}$ for a pair of

points a_1, $a_2 \in A$ (and we take $a_3 = a_2$ so that $x \in$ conv $\{a_1, a_2, a_3\}$). For, if no point of L lies in A then H is a convex set containing A but not x, which implies the contradiction $x \in \text{conv } A \subset \text{conv } H = H$, since $x \notin H$. Thus, there must exist $a \in A \cap L$. Consider the two opposing rays on L with origin x, say R_1 and R_2, with $a \in R_1$. Now $A \cap R_2 \neq \phi$ for otherwise $H \cup R_1 \setminus \{x\}$ is a convex set containing A but not x. Let $a_k \in A \cap R_k$, $k = 1, 2$, be such that $\rho(x,a_k)$ is minimal (assuming for the moment that A is closed; the argument for the case when A is not closed is very similar). If $\rho(x,a_1) + \rho(x,a_2) \geq \pi$ then $x \notin \text{conv}\{a_1, a_2\}$ and $\text{conv}\{a_1, a_2\}$ contains all the points of $A \cap L$ (by the minimal property of $\rho(x,a_k)$, $k = 1,2$); then one would have the convex set $H \cup \text{conv}\{a_1, a_2\}$ containing A but not x, a contradiction. Hence $\rho(x,a_1) + \rho(x,a_2) < \pi$ and it follows that $x \in \overline{a_1 a_2} = \text{conv}\{a_1, a_2\}$. This completes the case $x \in$ bd conv A.

If $x \in$ int conv A then let D be a disk centered at x with $D \subset$ conv A. There must exist $a_1 \in A$ such that a_1 lies in the open hemisphere H whose closure contains A, for otherwise A would be contained in the line L which determines H and conv $A \subset L$ or $D \not\subset$ conv A, a contradiction. Hence $a_1 \in A \cap H$ and we consider line $\overleftrightarrow{xa_1}$ and the two sides H_1 and H_2 of $\overleftrightarrow{xa_1}$. Certainly $A \cap H_k \neq \phi$ for $k = 1,2$, or else conv $A \subset \text{conv}(A \cap \overline{H}_{k+1})$ and $D \not\subset$ conv A. The closure of $A \cap H_k$ (the set $A \cap H_k$ and all its boundary points, denoted $\text{cl}(A \cap H_k)$) is compact, so there can be found

$a_2' \in cl(A \cap H_1)$ and $a_3' \in cl(A \cap H_2)$ such that $\sphericalangle a_1xa_2'$ and $\sphericalangle a_1xa_3'$ are maximal (Figure A.9). If $\sphericalangle a_1xa_2' + \sphericalangle a_1xa_3' \leq \pi$ then all of A would lie on one side of a line through x and we would have $D \not\subset$ conv A. Hence $\sphericalangle a_1xa_2' + \sphericalangle a_1xa_3' > \pi$ and there accordingly exist $a_2 \in A \cap H_1$, $a_3 \in A \cap H_2$ such that $\sphericalangle a_1xa_2 + \sphericalangle a_1xa_3 > \pi$ (Figure A.9). It follows since

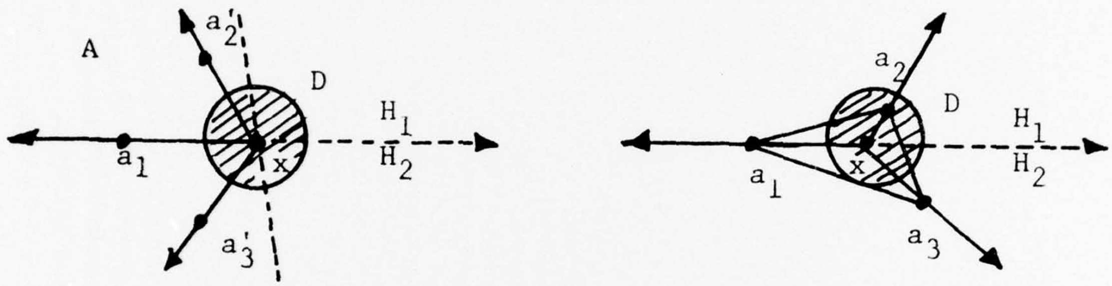


Figure A.9.

$a_1 \in H$ and a_2, $a_3 \in H \cap L$, that x is an interior point of triangle $a_1a_2a_3$. This implies that $x \in \text{conv}\{a_1, a_2, a_3\}$, finishing the case when A lies in a closed hemisphere. Finally, if A does not belong to a closed hemisphere then let H be a closed hemisphere which contains x, and define the set $A' = A \cap H$. Then

$$\text{conv } A' = \text{conv}(A \cap H) = \text{conv } A \cap \text{conv } H = (\text{conv } A) \cap H$$

and since x belongs to both conv A and H we have $x \in \text{conv } A'$ where A' is contained by a closed hemisphere. By the preceding case, $x \in \text{conv}\{a_1, a_2, a_3\}$ where a_1, a_2, a_3 belong to A', and therefore, also to A.

A.4 Convex Polytopes on S^2

A convex polytope is the convex hull of a finite set of points. As in classical convexity, every convex polytope in S^2 is compact. The question of when conv A is compact is nontrivial even though on S^2 any closed set is compact. Just as in the plane, there exist closed sets A such that conv A is not closed: Take a pair of antipodal points a, a' and a circle C through a' of radius less than $\pi/2$ (Figure A.10). Then take $A = a \cup C$; the convex hull of A is clearly the open hemisphere containing C, together with a and a'--not a closed set.

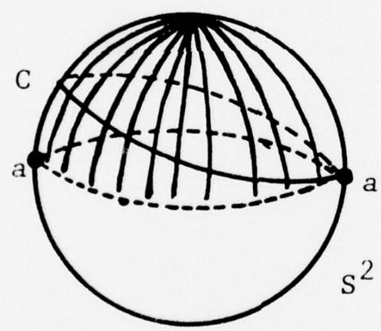


Figure A.10.

It is now shown that if A is finite then conv A is closed (therefore compact).

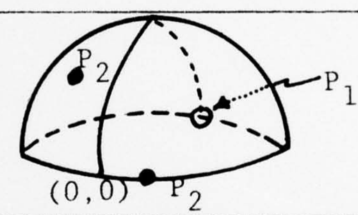
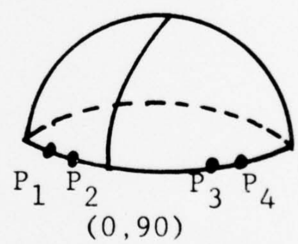
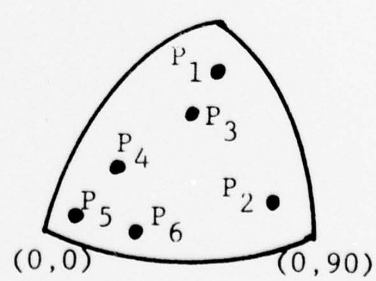
Lemma A.8 Any convex polytope on S^2 is compact.

Proof: Let $A = \{a_1, \ldots, a_n\}$, and consider a boundary point x of conv A; it is shown that $x \in$ conv A, proving that conv A contains its boundary points. For each integer

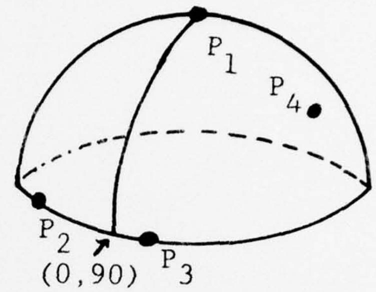
$k = 1, 2, \ldots$ let D_k be the disk centered at x having radius $1/k$. Then D_k contains a point $x_k \in \text{conv } A$, and $\lim_{k\to\infty} x_k = x$ (since $\lim_{k\to\infty} \rho(x, x_k) = 0$). Since S^2 has Carathéodory number 3 (Lemma A.7) there exist for each k a set of 3 points a_{1_k}, a_{2_k}, and a_{3_k} from $a_1, \ldots, a_n$ such that $x_k \in \text{conv}$ $\{a_{1_k}, a_{2_k}, a_{3_k}\}$. Since A is finite there is a subsequence $\{k'\}$ of k's for which $a_{1_{k'}}$ is constantly equal to a_u in A. Hence $x_{k'} \in \text{conv}\{a_u, a_{2_{k'}}, a_{3_{k'}}\}$. Again, a subsequence $\{k''\}$ of $\{k'\}$ exists such that $a_{2_{k''}} = a_v \in A$ and, finally, a subsequence $\{i\}$ of $\{k''\}$ such that $a_{3_i} = a_w$. Hence, a subsequence $\{i\}$ of $\{k\}$ exists such that $x_i \in \text{conv}\{a_u, a_v, a_w\}$, with $\lim_{i\to\infty} x_i = x$. Since conv $\{a_u, a_v, a_w\}$ is either a point, a segment, a line, or a nondegenerate triangle and interior (therefore closed), it follows that $x \in$ $\text{conv}\{a_u, a_v, a_w\} \subset \text{conv } A$, as was to have been proved.

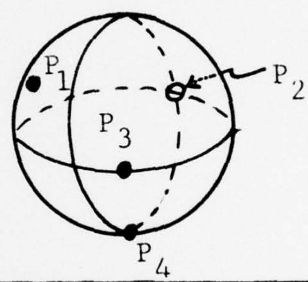
APPENDIX B

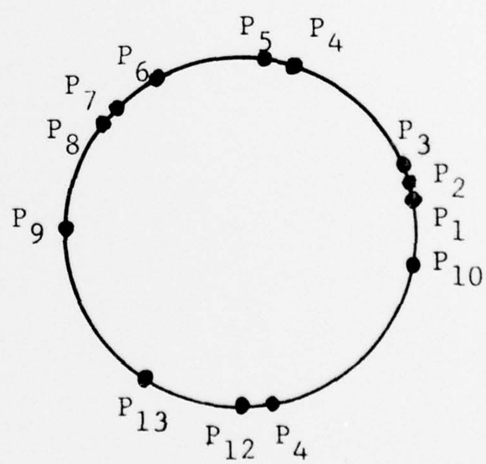
DATA FOR EXAMPLE PROBLEMS

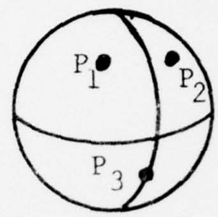
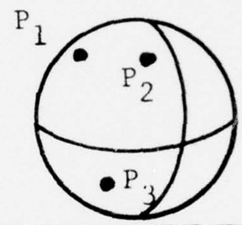
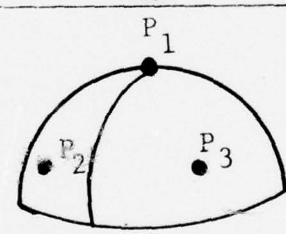
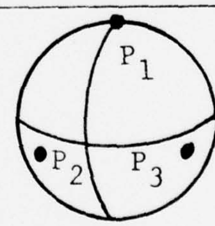
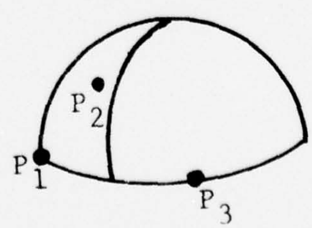
Number	Demand Point	Latitude Longitude in Degrees	Weight	
D1 See Figure 3.1	P_1	(0,180)	1	
	P_2	(0,36.86)	1	
	P_3	(30,-60)	1	
D2 See Figure 3.2	P_1	(0,40)	1	
	P_2	(0,55)	1	
	P_3	(0,145)	1	
	P_4	(0,160)	1	
D3 See Figure 3.6	P_1	(70,70)	1	
	P_2	(30,80)	1	
	P_3	(50,60)	1	
	P_4	(30,20)	5	
	P_5	(20,10)	5	
	P_6	(10,30)	5	

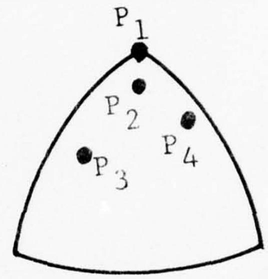
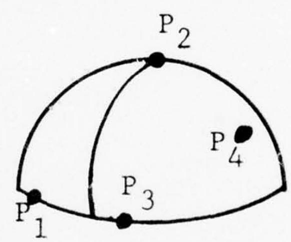
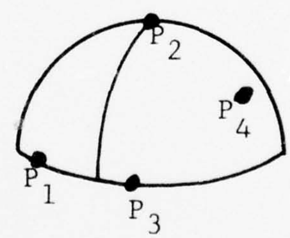
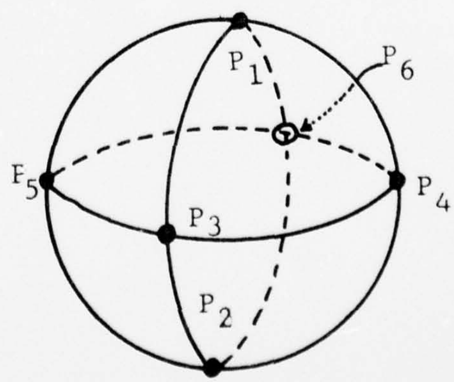
Number	Demand Point	Latitude Longitude in Degrees	Weight
D4 See Figure 3.10	P_1	(90,0)	1
	P_2	(0,20)	2
	P_3	(0,100)	1
	P_4	(40,160)	1
D5 See Figure 4.1	P_1	(30,-60)	1
	P_2	(0,180)	1
	P_3	(0,40)	1
	P_4	(-90,0)	1
D6 See Figure 5.1	P_1	(0,10)	1
	P_2	(0,16)	1
	P_3	(0,20)	1
	P_4	(0,70)	1
	P_5	(0,80)	1
	P_6	(0,120)	1
	P_7	(0,136)	1
	P_8	(0,140)	1
	P_9	(0,180)	1
	P_{10}	(0,-10)	1
	P_{11}	(0,-80)	1
	P_{12}	(0,-90)	1
	P_{13}	(0,-124)	1



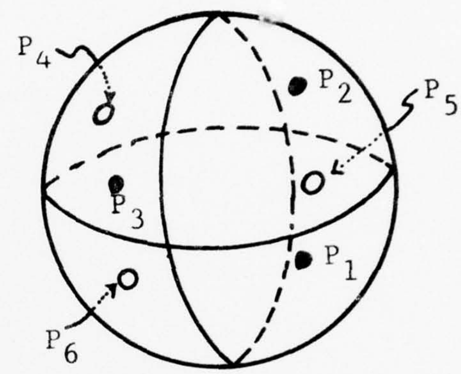


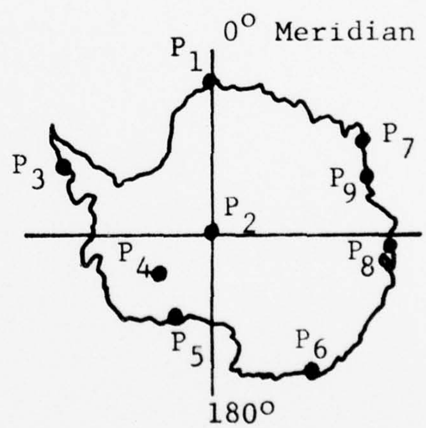


Number	Demand Point	Latitude Longitude in Degrees	Weight	
	P_1	(60,-30)	1	
D7	P_2	(60,30)	1	
	P_3	(-60,0)	1	
	P_1	(60,-90)	1	
D8	P_2	(60,-30)	1	
	P_3	(-60,-60)	1	
	P_1	(90,0)	1	
D9	P_2	(30,20)	1	
	P_3	(30,130)	1	
	P_1	(90,0)	1	
D10	P_2	(-30,20)	1	
	P_3	(-30,160)	1	
	P_1	(0,0)	1	
D11	P_2	(50,80)	1	
	P_3	(0,130)	1	

Number	Demand Point	Latitude Longitude in Degrees	Weight	
D12	P_1	(90,0)	1	
	P_2	(80,45)	1	
	P_3	(40,10)	1	
	P_4	(60,80)	1	
D13	P_1	(0,20)	1	
	P_2	(90,0)	1	
	P_3	(0,100)	1	
	P_4	(40,160)	2	
D14	P_1	(0,20)	1	
	P_2	(90,0)	1	
	P_3	(0,100)	1	
	P_4	(40,160)	1	
D15	P_1	(90,0)	1	
	P_2	(-90,0)	1	
	P_3	(0,0)	1	
	P_4	(0,90)	1	
	P_5	(0,-90)	1	
	P_6	(0,180)	1	

Number	Demand Point	Latitude Longitude in Degrees	Weight
D16 See Figure 5.2	P_1	(-12,48)	1.5
	P_2	(65,75)	3
	P_3	(15,-20)	2.5
	P_4	(25,-115)	2
	P_5	(-30,175)	3
	P_6	(-70,-110)	2
D17	P_1-Norway Station	(-70,0)	1
	P_2-Amundsen-Scott Stn.	(-90,0)	1
	P_3-Palmer Station	(-65,-65)	1
	P_4-Byrd Station	(-80,-120)	1
	P_5-Little America	(-78,-160)	1
	P_6-Port Martin	(-66,140)	1
	P_7-Mawson	(-68,62)	1
	P_8-Mirnyy	(-66,94)	1
	P_9-American Highland	(-70,75)	1





Number D18:

City	Latitude Longitude	Weight	City	Latitude Longitude	Weight
Minsk	(54,28)	0.012	Zhdanov	(47,38)	0.007
Lvov	(50,24)	0.010	Stalino	(48,37)	0.016
Kishinev	(47,28)	0.005	Makeyevka	(48,38)	0.008
Odessa	(46,31)	0.014	Gorlovka	(48,38)	0.007
Nikolaev	(47,32)	0.005	Taganrog	(47,39)	0.005
Kherson	(47,33)	0.004	Krasnodar	(45,39)	0.007
Sevastapol	(45,34)	0.004	Rostov	(47,40)	0.014
Simferopol	(45,34)	0.004	Shakhty	(48,40)	0.004
Krivoi Rog	(46,34)	0.009	Kadiyevka	(49,39)	0.004
Dneprodzerhinsk	(48,35)	0.004	Kharkhov	(41,37)	0.021
Dnepropetrovsk	(48,35)	0.015	Kiev	(50,31)	0.026
Zaporozhe	(48,36)	0.010	Gomel	(52,31)	0.004

Number D19:

Demand Point	City	Latitude Longitude	Weight
1	Montgomery, Alabama	$(32^{\circ}23', -86^{\circ}17')$	62,650
2	Juneau, Alaska	$(58^{\circ}25', -134^{\circ}30')$	1,163
3	Phoenix, Arizona	$(33^{\circ}30', -112^{\circ}00')$	22,288
4	Little Rock, Arkansas	$(34^{\circ}42', -92^{\circ}16')$	20,916
5	Sacramento, California	$(38^{\circ}35', -121^{\circ}30')$	394,139
6	Denver, Colorado	$(39^{\circ}44', -104^{\circ}59')$	47,453
7	Hartford, Connecticut	$(41^{\circ}45', -72^{\circ}40')$	74,813
8	Dover, Delaware	$(39^{\circ}10', -75^{\circ}30')$	15,863
9	Washington, D.C.	$(38^{\circ}50', -77^{\circ}00')$	238,476
10	Tallahassee, Florida	$(30^{\circ}25', -84^{\circ}17')$	80,791
11	Atlanta, Georgia	$(33^{\circ}45', -84^{\circ}23')$	79,198
12	Boise, Idaho	$(43^{\circ}38', -116^{\circ}12')$	9,502
13	Springfield, Illinois	$(39^{\circ}46', -89^{\circ}37')$	387,961
14	Indianapolis, Indiana	$(39^{\circ}45', -86^{\circ}08')$	91,294
15	Des Moines, Iowa	$(41^{\circ}35', -93^{\circ}37')$	60,204

Demand Point	City	Latitude Longitude	Weight
16	Topeka, Kansas	$(39^{o}02', -95^{o}41')$	47,086
17	Frankfort, Kentucky	$(38^{o}10', -84^{o}55')$	41,344
18	Baton Rouge, Louisiana	$(30^{o}28', -91^{o}10')$	52,978
19	Augusta, Maine	$(44^{o}19', -69^{o}42')$	17,631
20	Annapolis, Maryland	$(39^{o}00', -76^{o}25')$	73,834
21	Boston, Massachusetts	$(42^{o}15', -71^{o}07')$	153,409
22	Lansing, Michigan	$(42^{o}45', -84^{o}35')$	127,187
23	St. Paul, Minnesota	$(44^{o}57', -93^{o}05')$	90,323
24	Jackson, Mississippi	$(32^{o}17', -90^{o}10')$	24,448
25	Jefferson City, Missouri	$(38^{o}34', -92^{o}10')$	139,140
26	Helena, Montana	$(46^{o}34', -112^{o}01')$	17,322
27	Lincoln, Nebraksa	$(40^{o}49', -96^{o}43')$	35,658
28	Carson City, Nevada	$(39^{o}10', -119^{o}45')$	9,207
29	Concord, New Hampshire	$(43^{o}10', -71^{o}30')$	11,631
30	Trenton, New Jersey	$(40^{o}13', -74^{o}46')$	184,397
31	Sante Fe, New Mexico	$(35^{o}10', -106^{o}00')$	17,645
32	Albany, New York	$(42^{o}40', -73^{o}50')$	662,584
33	Raleigh, North Carolina	$(35^{o}45', -78^{o}39')$	73,749
34	Bismark, North Dakota	$(46^{o}48', -100^{o}46')$	11,646
35	Columbus, Ohio	$(40^{o}00', -83^{o}00')$	219,330
36	Oklahoma City, Oklahoma	$(35^{o}27', -97^{o}32')$	59,159
37	Salem, Oregon	$(44^{o}55', -123^{o}03')$	45,733
38	Harrisburg, Pennsylvania	$(40^{o}15', -76^{o}50')$	302,933
39	Providence, Rhode Island	$(41^{o}50', -71^{o}23')$	22,769
40	Columbia, South Carolina	$(34^{o}00', -81^{o}00')$	28,434
41	Pierre, South Dakota	$(44^{o}22', -100^{o}20')$	10,292
42	Nashville, Tennessee	$(36^{o}10', -86^{o}48')$	68,770
43	Austin, Texas	$(30^{o}15', -97^{o}42')$	233,041
44	Salt Lake City, Utah	$(40^{o}45', -111^{o}52')$	23,110
45	Montpelier, Vermont	$(44^{o}20', -72^{o}35')$	14,082
46	Richmond, Virginia	$(37^{o}35', -77^{o}30')$	74,408

Demand Point	City	Latitude Longitude	Weight
47	Olympia, Washington	$(47^{o}02', -122^{o}52')$	53,472
48	Charleston, West Virginia	$(38^{o}20', -81^{o}35')$	23,240
49	Madison, Wisconsin	$(43^{o}05', -89^{o}23')$	86,544
50	Cheyenne, Wyoming	$(41^{o}10', -104^{o}49')$	7,489

VITA

Daniel W Litwhiler, Jr. was born on February 28, 1942 in Ringtown, Pennsylvania. In 1960 he graduated from Florida High School in Tallahassee, Florida.

In 1963 he received a dual Bachelor of Science degree in mathematics and mathematics education (with honors) from Florida State University.

In 1965 he received the Master of Science degree with a major in mathematics from Florida State University and entered the United States Air Force as a career officer.

He held a teaching assistantship in the Department of Mathematics at Florida State University from 1963 to 1965, and taught extension courses in mathematics for the University of Southern Mississippi in Biloxi, Mississippi, in 1966.

Since entering the Air Force he has held a variety of positions including Squadron Operations Officer for a remote "listening post" site, Comptroller for Southern Communications Area (AFCS), Programs Officer for Fifth Air Force responsible for monitoring all Air Force aspects of the Okinawa Reversion, and Assistant Professor in the USAF Academy Department of Mathematical Sciences. He is currently at the Air Force Academy.

Major Litwhiler is a member of Phi Eta Sigma, Pi Mu Epsilon, AIIE, and ORSA.